DOCUMENT RESUME

ED 081 007 CS 200 665

AUTHOR	Purves, Alan C., Ed.
TITLE	1972 Minnesota NCTE (National Council of Teachers of English) Seminar on Research in English Education.
INSTITUTION	National Council of Teachers of English, Urbana, Ill.
PUB DATE	73
NOTE	150p.; Papers presented at the Invitational Seminar on Research Design in English Education (Minneapolis, November 1972)
AVAILABLE FROM	National Council of Teachers of English, 1111 Kenyon Road, Urbana, Ill. 61801 (Stock No. 98566-7310, $1.75 non-member, $1.50 member)
JOURNAL CIT	Research in the Teaching of English; v7 n2 p145-290 Fall 1973
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Anthropology; *Conference Reports; *Educational Research; English Education; *English Instruction; Literature; *Literature Reviews; *Research Reviews (Publications); Testing
IDENTIFIERS	NCTE Seminar on Research in English Education

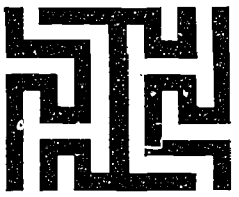
ABSTRACT

The articles in this journal relate the conference experience at the 1972 Minnesota National Council of Teachers of English (NCTE) Seminar on Research in English Education. Articles include "The New Research" by Peter S Rosenbaum (discussing the influence of the seminar experience and how it can precipitate new research); "Research in the Teaching of English: The Troubled Dream" by Dwight L. Burton (commenting on the need for quality research in English and the problems involved); "Anthropological Research Models" by Carol Talbert; "Measurement and Evaluation" by Jeremy D. Finn; "Research into Imagic Association and Cognitive Interpretation" by Harry S. Broudy; "Some Types of Research on Response to Literature" by Gunnar Hansson; and "An Experience Preserved: The Video Tapes of the Minnesota-NCTE Seminar" by Michael F. Graves. (HOD)

1972 Minnesota NCTE (National Council of Teachers of English) Seminar on Research in English Education

Alan C. Purves
Editor

Fall 1973



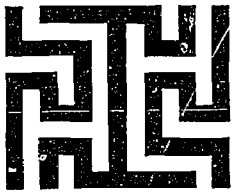
RESEARCH IN THE TEACHING OF ENGLISH

Alan C. Purves, *Editor*
University of Illinois
310 West Delaware
Urbana, Illinois 61801

James Hoetker, *Associate Editor*
College of Education
Florida State University
Tallahassee, Fla. 32306

Contributors should follow the format of *Research in the Teaching of English*. Articles should include captions at appropriate places. Articles should be submitted in duplicate with a self-addressed return envelope (to which stamps are clipped) to Alan C. Purves, 310 W. Delaware, Urbana, Illinois 61801. Deadlines for manuscripts: November 1 for spring issue, May 1 for fall issue.

Published winter, spring, and fall by the National Council of Teachers of English. Subscription price, $5.00 per year; single copies, $1.50 each. Send subscription and business communications to NCTE, 1111 Kenyon Road, Urbana, Illinois 61801. Entered as second-class matter January 31, 1967, at the Post Office at Champaign, Illinois 61820, under the Act of March 3, 1879. Printed by R. F. Colwell Printing Corporation, 510 North Hickory Street, Champaign, Illinois 61820.



RESEARCH IN THE
TEACHING OF ENGLISH

VOLUME 7, NUMBER 2, FALL 1973

Contents

ARTICLES

Editor's Foreword

It is with great pleasure that *Research in the Teaching of English* inaugurates its new schedule of three issues per year with a special issue devoted to the 1972 Minnesota NCTE Seminar on Research in English Education. As Professor Rosenbaum indicates in his summary article, the Seminar was a major event for those who participated. The following articles and reports will capture a part of its flavor; the rest must come from a viewing of the videotapes described by Professor Graves.

A special issue initiates the new schedule of publication in a fitting manner because we have found that *RTE* was bursting at the seams. The number of worthy articles submitted for publication has grown. The bibliography has expanded. The need for reviews and articles of news has increased. A thrice-yearly schedule—fall, winter, and spring—seemed an appropriate step. We are grateful to the Trustees of the Research Foundation of the National Council of Teachers of English for their support of this expansion and of the issue now before you.

This special issue is somewhat out of phase with what we hope will be the normal production schedule. The Fall and Winter issues will be devoted to submitted reports of research and articles on research topics. The semi-annual bibliography will also appear in these issues. The Spring issue will be devoted to reviews and articles dealing with broader conceptual articles related to research in the teaching of English. That issue will be edited by Professor Hoetker. The first review issue will appear in 1974.

The New Research

PETER S. ROSENBAUM
Teachers College, Columbia University

If the gathering cited in the headnote of this essay endures in the memory of the profession, it is most unlikely to do so for any cause that could be directly attributed to the papers that follow in this volume, which, taken as a collection, constitute the information base to which the NCTE conference applied its energies. This opinion is offered with no intended disrespect either for the quality or for the insightfulness of these works; the point is rather this: their content possesses no special characteristics that would identify them with the conference. Each paper stands alone, self-contained. Without explicit conference identification, the reader could conclude his dealings with them without ever realizing that a conference has taken place. To the extent that this particular NCTE conference will emerge as a landmark, it will do so less because of its tangible output than because of its impact *as an experience* on the minds of those people who participated in it. To a certain degree, a researcher's thinking is altered, however minutely, in every professional encounter. But for many participants in this conference, the experience occasioned such a scrambling of basic assumptions regarding research in English Education as to augur a profound alteration in the structure of existing patterns of research activity. "Research Design," as a label, exists still; the NCTE conference did not record any change in the name of the game. But it did witness, however briefly, and with degrees of effect varying from individual to individual, a change in the game of the name. And in whatever way that the conference experience has compelled its participants to think and act in new ways, and to influence colleagues, students, and clients

in the process, in that way will it become worthy of recollection.

If the conjecture developed in this report is correct, the Minnesota NCTE Seminar has breathed life into a New Research.

THE INITIAL UNSETTLING

The first suggestion that there could be anything unusual about this conference emerges in the juxtaposition of the program content of the conference with the prior expectations of the participants, for whom it was a special event. After all, conferences of this sort are infrequently sponsored by NCTE, primarily because of the associated expense. Further, the conference was invitational, giving it an aura of exclusivity and high purpose; and too, it was carefully and elaborately planned over a period of more than a year. What confusions and uncertainties are thus occasioned during one's first encounter with the titles of the papers that were to comprise the program content! Considered individually, as journal offerings perhaps, these titles are without fault, but as the agenda for a summit conference on Research in English Education, well, at the very least, one would feel compelled to ask a few questions. If "anthropological research models," why then not sociological models, psychological models, ecological models?

MINNESOTA NCTE SEMINAR
ON RESEARCH IN ENGLISH EDUCATION

Invited Papers:

"Research in the Teaching of English: The Troubled Dream"

"Anthropological Research Models"

"Measurement and Evaluation"

"Research into Imagic Association and Cognitive Interpretation"

"Some Types of Research on Response to Literature"

"A Dynamic Medium for Creative Thought"[1]

[1] Unfortunately "A Dynamic Medium for Creative Thought" by Alan Kay is not printed in this volume for technical reasons. Much of the paper was a slide-tape demonstration not suitable for reproduction in print.

If "response to literature," why then not research on creative and expository writing? And what connection is supposed to be drawn between the notion of research and "a dynamic medium for creative thought"? And with what strand of research in English Education is "imagic association and cognitive interpretation" to be identified? The mind stumbles when it attempts to reconcile this selection of topics with its expectations. The list seems loose and unorganized, not at all what one might have anticipated.

And then, all at once and for the first time, the mind finds itself entertaining the possibility that the problem may not reside in the program content at all, but rather with the expectations, in other words, that the mind has been hypnotized by its own preconception of what research is. In defense, the mind retreats into a state of suspended judgment. For, yes, it is still possible that the content of the conference is defective; but it is equally possible that the program content is perfect and that the mind has thus far failed to get the message: Research in English Education is not a "babbling brook."

A SECOND ENCOUNTER WITH THE PROGRAM

In retrospect, it was the *roles* to be played by the authors of the conference papers during the conference, rather than the subjects of their papers, that are most suggestive of a hidden agenda.

MINNESOTA NCTE SEMINARS
ON RESEARCH IN ENGLISH EDUCATION

Invited Roles:

- a researcher laureate in English Education
- an anthropologist
- a psychometrician
- a philosopher
- a European researcher
- a technologist

The diversity of role is so striking here, the departure from homogeneity so complete, as to suggest premeditation, deliberate intent. Inquires the mind, Why the diversity? For its own sake? Perhaps, but unlikely, for the invitations do seem clearly to reflect a relatedness to English Education. More likely, this explicit diversity is in-

tended to convey and support the fact that the very definitions of research are not absolute definitions, but definitions *relative* to specific points of view, the philosopher's, the anthropologist's, the technologist's. The past quarter century, with its intensification of human interaction through electronic mediation and communication, has witnessed the fall of one constant after another from the heights of exclusivity and timelessness. Has the moment not come (perhaps even passed) for Research in English Education? Perhaps the conference agenda is telling us that a new richness of understanding will emerge in the process of coming to grips with growing uncertainty, that ideas and positions can no longer be worked out in the context of the refinement or deepening of a specific and established frame of reference, that what is called for is rather a magical feat of *extension* in which the mind permits itself to play over all available perspectives, seeking new linkages, building new configurations.

If all this is plausible, the content organization finds, at last, a reasonable interpretation, for the invited speakers, by virtue of their diversity of interest and outlook, could be expected to serve as invaluable catalysts and foils in a conference striving to lay bare existing overlaps among nominally distinct frames of reference. It seems not to have been the message of the conference that we throw away our intellectual technology, but rather that we step outside of ourselves in order to understand the biases of existing conceptions of research. Thus, the selection of contributors comes very close to the ideal, for what a rich galaxy of thought this group could be expected to generate!

SEMINAR DESIGN

If extension, interconnection, and overlap formed the tacit agenda for the NCTE Seminar, this message was confirmed, reenforced, and amplified by the very design of the conference. If the agenda was the bait, the conference design was the beaters. There was no escape.

The *physical environment* for the conference was a small motel in St. Paul, an unwieldy distance from the metropolitan attractions of the Twin Cities. Diversity had to be dealt with on a very practical level, for if not in the isolation of his room, a participant was inevitably in contact with other participants.

The *size of the conference*, too, encouraged mutuality. The list of participants numbered no more than sixty, thus fostering an intimacy with respect to the Seminar that probably could not have been achieved in more spacious facilities with a greater number of participants. In a more commodious setting, one could easily have found refuge in a mass identity; upon encounter with a colleague, one would likely not feel compelled to engage in discussion of the issues of the conference. But at this conference, interaction seemed to acquire a sense of urgency and importance. In a sense, the participants were sharing a lifeboat with their colleagues. One saw the same people, over and over again, and as the drama of the conference mounted, more and more would a participant be inclined to seek counsel and discussion with other participants, not necessarily because of a perceived convergence of professional interest, but simply because other participants *were there* and could be expected to have opinions about the proceedings.

Then too, there was the *treatment of literal content*. The papers that follow in this volume were commissioned by NCTE's Committee on Research not for presentation at the Seminar, but rather to serve as a content base for discussion. The papers were distributed in advance; no event took place that featured the direct transmittal of information. The conference was organized around a series of telerecorded panel discussions, each involving the author of one of the papers, a moderator, and two or three consultants selected in advance from the participant list apparently according to criteria again suggestive of the desire to establish overlap, a linking up of diverse points of view. And then there was the "studio audience," the participants themselves, watching the show live, as it were, contributing to it and influencing its direction, as a Greek chorus, as a jury, as a sounding board.

What mysterious effects accompanied the television aspect of the conference! Perhaps more because of the television than any other single factor, the content of the conference began to emerge in the mind not as the papers under discussion, but as the participants themselves. Each of us was the content, our opinions, our responses, our actions. It was in these televised sessions that the mind could first

begin to discern that the definitional foundations of research were being loosened, were being shaken free from their anchors in the visual and graphic world of the professional literature. Wherever the issues lay, they were being sought not through the once-removed world of print, but through human encounter in a family setting, through direct, first-hand confrontation in real time.

Somehow, the television converted the meeting into *a family affair,* fostering a spirit of cooperation and mutual involvement, and engendering a commitment to seek resolution of issues communally. No one who participated in the NCTE Seminar will ever forget the incident of interchange during one of the televised panel discussions where an apparent inability of the participants to achieve understanding resulted in the rejoinder remembered as "Well, you do your thing and I'll do mine"—a total breakdown in communication, and a jarring, discordant note in a setting that had been designed not merely to be supportive of tolerance for different approaches, but to achieve a predisposition to seek new syntheses based upon different approaches.

Working Parties

As if the televised "Plenary Sessions" had not established sufficient basis for overlap to become perceptible, the participants had been preassigned to one of six "Working Parties" that nominally were to reflect distinct disciplinary perspectives.

Working Party Designations
Language Development
Reading
Interdisciplinary Studies
Sociolinguistics, Psycholinguistics, and Reading
Literature
Composing and Speech
Method and Curriculum

The Working Parties met after each televised panel discussion to contemplate the content of these discussions in the light of the designated disciplinary identifications or,

at the very least, in the light of whatever perspectives would actually emerge in the course of Working Party deliberations. As it turned out, homogeneity of outlook was achieved rather imperfectly, several parties seeming to be more microcosmic of the entire conference than of specific disciplinary interests—not that it seemed to matter in the least.

Thus, in the beginning there were the conference papers, delivered to the participants and read by them in advance. Second, there was a new amalgam generated by the conference itself in the context of the televised panel discussions. Finally, there was the dissection and reconstitution of these data in the smaller groups, which allowed greater freedom of expression to the participants and more time to explore in depth. What a scrambling, what a dismantling, what confusion, what new riches, for the totality of a new mosaic was there for all to see and experience.

Sequence

And, finally, the keystone of the Seminar design, the impelling force, a task- and goal-oriented *sequence of events* that had been conceived in such a manner as to guarantee the inevitability of closure. Following the last Working Party session devoted to a panel discussion, the "recorders" of each Working Party were asked to develop and submit summary statements of their deliberations throughout the conference, these statements to be duplicated overnight and distributed to all of the Working Parties the next day so that each group might generate a final synthesis of understanding based upon consideration of the thoughts of all of the Working Parties. There would then occur a final televised Plenary Session during which the results of the closure process would be scrutinized and elevated to public view. In short, there was no escape; we all shared in the challenge, and it was sink or swim. We felt that we had a job to do and that we had a mutual responsibility toward one another. And, so, we joined together and did the job, changing ourselves in the process.

The design aspects of the conference, the conference *medium*, if you will, bear not merely upon the question of what Research in English Education is, but even more fundamentally upon a new vision of how to approach

research and its associated priorities. Under this vision, truth is perceptible not with the full-blown all-at-onceness that the print medium leads us to anticipate. Rather, truth reveals itself tentatively, gradually, fleetingly, through a *scanning* process, with all of the uncertainty and promise of blips on a radar screen. An idea comes to be sensed as important not because of its immediacy, but because the researcher experiences a continual reemergence of the idea over time, as with the recurrent themes of the NCTE Seminar. An idea appears on the screen and keeps reappearing with each sweep of the luminescent scanner; it will not remain forever, but the longer it remains and the larger it becomes, the more prominent becomes its presence in the mind, perhaps as an idea, perhaps as an issue, perhaps as a priority, perhaps as an intuition. Research becomes a dynamic, changing affair, achieving, one might imagine, or at least hope, an easier, more comfortable relationship with the world of data and events, a Research Design for exploring the world, yes, but being a part of it at the same time.

THE BLIPS ON THE SCREEN

The medium of a "report" inexorably imposes on the conscience of the writer the obligation to deal with the content of events as well as with the events themselves, notwithstanding the fact that without its experiential aspect the content of the NCTE Seminar is as dry and sparse as the physics of a golf swing. To the extent that the content of the NCTE Seminar was revealed to the participants, it was revealed not through print, but through experience, through scanning, through the recurrence of themes in informal meetings, in the Plenary Sessions, and in the meetings of the Working Parties. To the extent that the Seminar is revealed to others through the viewing and study of videotaped record of plenary sessions, produced by Michael Graves, it may seem fragmentary, diffuse, and highly varied. It is perhaps best here, therefore, to make no attempt at a literal rendering, but rather to attempt a reconstruction of the blips that did, in fact, appear on the screen for those three categories of events where documentation is extant. Perhaps in this way the reader will come to sense the rhythm and shape of issues and concerns.

The information basis of the NCTE Seminar as embodied in the invited Seminar papers.

BLIP

A proposed style of research premised upon a liberating notion that would establish goals not in terms of population norms (i.e., "the average student"), but in terms of the actual observed differences in people with various levels of formal schooling; and a strategy of schooling that would replace the literal by the interpretive.

BLIP

A theory of evaluation that relates the technology of evaluation to practical bases upon which quality rationales can be constructed.

BLIP

A new medium (electronic) that could make graphic design an inexpensively accessible public medium and art form, with the likelihood of profound impact on the foundations of the English curriculum.

BLIP

Notes of hesitancy, uncertainty, discontinuity, questioning, seeking, suggesting. . . .

BLIP

An anthropological overlay on educational research, stressing fewer statistics and more personal involvement, fewer rigid conceptual forms for investigating educational processes, and greater willingness to consider cultural factors that influence behavior in educational settings.

BLIP

An investigation into a human being's internal responses to information rendered in the literary print medium, and the perhaps-to-be-expected difficulties of exploring in such a private domain.

The Golden Moments of Television (as summarized from listings prepared by each Working Party during their final meeting).

BLIP

Harry Broudy's eloquence in pleading for sense in the discussion of intended classroom outcomes and for a broadening of our view of behavior relative to educational goals, but equally his personal control, a model performance.

BLIP

The friction between Carol Talbert and Robert Calfee on the matter of the perspective of the researcher, on matters of "rightness," "absoluteness," and "relativity"; and the breakdown of the interchange between the two, a moment of drama.

BLIP

Janet Emig's gnawing and compelling amplification of Dwight Burton's opening remarks, specifically suggesting a reconceptualization of both the purposes and the scope of research in English Education and a broadening of both.

BLIP

The uncomfortable inability of Gunnar Hansson and James Britton to achieve overlap and closure on basic assumptions regarding literature and literary response.

BLIP

Julie Jensen's direct questioning of the relevance of Hansson's line of research, suggesting a broader scope of interest and elevating such a broadening to a general issue.

BLIP

Kenneth Johnson's witty, aggressive, and intimately informal underscoring of the point that any single perspective carries with it biases to which the investigator is blinded, that such biases become readily accessible

to objective view only when viewed from different perspectives; in a real sense, therefore, stating a recurrent theme that what is called for is not the deepening of existing perspectives, but the treatment of perspective as content.

BLIP

Alan Kay's description of new instructional modes growing out of computer technology.

BLIP

Peter Rosenbaum's understated observation that if one is seeking to understand the significance of media technology for English Education, one should not permit hardware to induce hypnosis, for it is as much the style of thought in which the development of media technology is couched as it is the hardware that is likely to result in curricular and pedagogical transformations.

BLIP

The stir created by James Hoetker's observation concerning the *de facto* influence of political and social factors upon research and Research Design.

Recurrent themes in the Working Party discussions (as quoted or adapted from summaries prepared during their penultimate sessions).

BLIP

"The nature of 'research' seemed to be in question throughout."

BLIP

"Great need for research centers."

BLIP

"Need for interdisciplinary understanding."

BLIP

"Reading research is at present fragmented and at a preparadigmatic stage. There is a need to synthesize basic concepts from existing reading studies."

BLIP

"Let the research be suited to the question."

BLIP

"The reward system (of the profession) discourages integration in research and encourages fragmentation. This must be changed."

BLIP

"The committee (i.e., Working Party) on Interdisciplinary Studies encourages researchers to employ the perspectives and methodologies of many disciplines (e.g., anthropology, sociology, psychology, linguistics, etc.). The types of research which utilize the strategies of many disciplines will most likely provide the best evidence possible in examining research questions."

BLIP

"English educators must be sufficiently trained in statistics to talk intelligently to statisticians. One course in statistics is not enough."

BLIP

"They also need enough knowledge of anthropology and sociology to talk to anthropologists and sociologists."

BLIP

"How are the responses to literature related to perceived goals of education in literature?"

BLIP

"Explore what a research center might be like."

BLIP

"Don't be apologetic about nonquantifiable data. Explore other methods—e.g., case studies, interaction analysis, etc."

BLIP

"We need above all to broaden our own notion of what research is."

BLIP

"Graduate students need to be involved in research throughout their schooling, not merely when engaging in dissertation study."

BLIP

"We need and must support more 'conceptual' research, based upon present knowledge, and leading to new theoretical frameworks."

BLIP

"Our research questions must be those that are asked in ways that are meaningful or applicable to school situations and teachers. We cannot afford to ignore the pressing real-world problems, and we cannot expect that some other middleman will use our findings for us."

THE MEANING OF THE SEMINAR

In the final analysis, from what perspective is the conference itself to be viewed, in what way does it acquire its meaning and deliver its message?

Perhaps one way of coming to grips with this question is to start with the recognition that the design idiosyncrasies of this particular NCTE conference were not fortuitous; they were deliberately planned by the NCTE Committee on Research under the leadership of James Squire.[1] It is thus not altogether too far-fetched to spec-

[1] The planning committee included Charles C. Cooper and Walter Petty, State University of New York, Buffalo; Thomas Devine, Boston University; Peter Rosenbaum, Teachers College, Columbia University, conference recorder; and James R. Squire. Doris Gunderson, Alan Purves, and Robert Hogan assisted in selection of participants. Michael Graves and Eugene Piché, University of Minnesota, were cochairmen for arrangements, and Mr. Graves produced the videotaped record.

ulate that content, structures, and processes of the conference, hence its message, evidence a Grand Scheme that originated in Squire's imagination. In all aspects that gave the conference its special flavor, Squire's hand can be seen; in the selection of the contributions, in the design of the communications structure of the conference, in the formating of events, and in the establishment of their sequence, and in the designation of physical setting. For sure, Squire's role was vital.

But if the ultimate source of the message is Squire, and if this conference was his way of communicating with the profession, there is an equally compelling sense in which Squire could be acting as the unwitting agent of a general societal trend of such magnitude and profundity as to alter fundamentally whatever it may touch, including Research in English Education. No matter what the impetus provided by Jim Squire and the Committee on Research, it is inconceivable that we could witness the events reported on herein if the participants themselves had not been prepared for such a happening. If Squire created the vehicle, we were all willing to go along for the ride; if Squire created the medium, we were all willing to take the plunge, to immerse ourselves in the new experience.

From the outset of the conference, the tone was one of toleration, of conciliation, almost as if the participants sensed something in the wind. Personal animosity, to the extent that it revealed itself at all, served not as a call to battle, but as an unhappy note of failure; dramatic, amusing, justified at times—perhaps, but always disconcerting, for it implied an inability of the participants to examine their own biases objectively. Tenseness and fence-building, as it may have existed during the early meetings of the Working Parties, gradually lessened as the incompatibility of such behavior with the drift and felt intent of the conference became apparent. Barriers were lowered and replaced by linkages of mutual commitment. Those in the profession who can remember the aggressive zeal with which argumentation used to rage in comparable professional gatherings not just a few years ago will pause here for reflection. The point is not that we have exhausted the supply of ideas to be studied and acted upon, but rather that our attitudes appear to be changing in such a

manner as to alter the structure of our professional interaction. Perhaps we are ridding ourselves of our disciplinary ethnicity.

It is worth noting that the general pattern of information processing at the conference directly reflects changes that have taken place in each of us, individually, as the result of our membership in an electronic society. We have, as others have noted, become information processors on a grand scale. Styles of research are likely to change simply as a consequence of the growing perceived necessity of conserving real time. Many readers of this volume will not "read" it at all, at least not in the sense of reading that is implicit in the tests used to measure reading achievement in schools. The New Reader will *scan* this volume; he will *program* his interaction with commercial television with a *flip* of the TV selector dial. The New Reading is a tool for dealing with a new reality. As McLuhan observes, men are not left untouched by the tools they invent and use. New tools make new men. The NCTE Seminar revealed to its participants, however fleetingly, however tentatively, however vaguely, a new conceptual tool, a new and compelling technique for seeing. And in the application of this new tool will arise the New Research, a research that will be global in its effect on how research is conceived, how it is executed, and how it segments and labels the reality to which it is applied.

Such is the message of 1972 Minnesota NCTE Seminar on Research in English Education.

We have thus a situation of unusual kinetic potential. The mind has been tantalized, but the shape and substance of what is to come, of actions and events that are to shape the directions of the profession, are yet hidden from our view and will take form only gradually in the course of our interaction with colleagues, students, and clients.

But surely, by the publication of the papers for the next NCTE Seminar on Research, the secret will be out.

Research in the Teaching of English: The Troubled Dream

DWIGHT L. BURTON
Florida State University

Why speak about research in the teaching of English as a "troubled dream," other than to inject a humanistic fillip into an otherwise pedestrian title, though "troubled dream," as far as the writer knows, is not a specific allusion to anything in Shakespeare, Wordsworth, or T. S. Eliot. "Research" has a highly positive connotation in American culture, suggesting the rational, scientific approach to truth and knowledge, and is the premium of academe, though in the humanistic-oriented English teaching profession there has been an abiding uneasiness with quantitative methods and perhaps with the empirical approach generally. Yet we have great hopes for what research can do for us in the teaching of English. Though at the moment, we acknowledge that research has had little to do with curriculum structure and teaching methods in English, we have the feeling that answers are just around the corner if we could but design the right studies. To do better, with appropriately more important implications for teaching, what we have done before is the objective of many involved with research in the field. J. Stephen Sherwin, for example, concludes his critique of 1969 on research in selected areas of the teaching of English with the caveat, "All we can do is to proceed along the path already taken and to try to step more cautiously and wisely. The urgent need is for improved techniques for educational research and for all

other behavioral investigations".[1] And Edward Summers, reviewing research in the teaching of reading in secondary schools in 1967, avers that new questions are not needed, only better research attacks on those already raised.[2] Thus, the dream part.

The troubled part refers to the disappointments and frustrations, consistently and almost monotonously voiced, with the fruits of research that so far has been done in the teaching of English. For instance, the authors of the "state of knowledge" reports so far published by the National Council of Teacher of English find much more to discuss about what research has not accomplished than about what it has.[3] Most reviewers of research shake their heads sadly over the quality of much of what they review, and listings of common faults that make research inconclusive or invalid are strikingly similar over the past twenty-five years.

One would hope that the dream would persevere and become more serene in the years ahead. Perhaps we have expected too much of an effort that is relatively young and we may have underestimated the difficulty of probing, for our particular reasons, the mysteries of human behavior, its development and change. Actually, the student of research in the teaching of English can find much to rejoice about as he looks back over the last decade. Research activity in the field has represented a virtual explosion. The commitment of the profession generally to research is far greater now than it was a few years ago. Federal funding, though now drastically and, hopefully temporarily reduced, has supported the effort on a large scale.

---

[1] J. S. Sherwin, *Four Problems in Teaching English: A Critique of Research* (Scranton, Penn.: International Textbook Co., 1969). p. 189.

[2] E. Summers, "Reading in the secondary school," *Review of Educational Research* XXXVII (April 1967), pp. 134–151.

[3] R. Braddock, R. Lloyd-Jones, and L. Schoer, *Research in Written Composition* (Champaign, Ill.: NCTE, 1963); W. Petty, P. Herold, and E. Stoll, *The State of Knowledge about the Teaching of Vocabulary* (Champaign, Ill.: NCTE, 1968); A. C. Purves and R. Beach, *Literature and the Reader: Research in Response to Literature, Reading Interests, and the Teaching of Literature* (Urbana, Ill.: NCTE, 1972).

A new journal, *Research in the Teaching of English*, began publication in 1967.

EMPHASES AND PROBLEMS OF RESEARCH IN RECENT YEARS

Substantive Emphases

What directions and emphases may show through the burgeoning of research in recent years? Since *Research in the Teaching of English* began publication, Nathan S. Blount has prepared for each semi-annual issue a bibliography of research in the field. His bibliography for the period January 1, 1971–June 30, 1971 lists 389 references, 196 of them in the teaching of reading. The next largest category (41 references) Blount labels "Verbal Learning and Cognitive Dvelopment." Categories with approximately equal numbers of references (about 25 each) are: 1. The English Teacher and the Curriculum—status surveys, methods and materials; 2. Teaching Literature; 3. Teaching English Skills; 4. Teaching Written Composition and Rhetoric; 5. Teaching Oral Language. The final two categories—Teacher Education and Teaching English as a Second Language—include about a dozen references each.

Of the eleven studies reported in the two issues of *RTE* for 1971, two each are in language development, teacher education, the teaching of literature, and the teaching of writing; one each are in the teaching of grammar, the teaching of reading comprehension, and the combined teaching of writing and reading.

Further light on substantive areas of emphasis in recent research is shed by those studies which have received particular recognition through the distinguished research awards, the promising researcher citations, and the research monographs of the National Council of Teachers of English.[4] The distinguished research awards are given either for individual studies or for general contributions in the field. Of the eight annual awards so far, all except one have been given for research in linguistics or language development (the exception being Wayne Booth's *The Rhetoric of Fiction*). Four of the awards were for studies

[4] This paper was prepared mainly in the early spring of 1972. The Distinguished research awards and the promising researcher citations are announced each year at the November meeting of the NCTE. The monographs are published on an irregular schedule.

in the language development of children and adolescents.[5] One award went to a study of the social stratification of English in New York City,[6] while two were given for general contributions in linguistic research to Raven I. McDavid and Albert H. Marckwardt.

The promising researcher citations instituted by the NCTE in 1970 are made for initial major research of exceptional merit. All of the twelve citations so far have been of doctoral dissertations in the general area of language development or the relation of language factors to reading comprehension or writing performance except one which was a study of the implications of Burke's rehtoric for the high school program.

Thirteen research monographs have been published by the NCTE as of this writing, several reporting the same studies which won distinguished research awards. Four deal with language development, three with response to literature, two with the effects of transformational grammar on writing performance, and one each with: NCTE attitudes toward language; the relationship of drama study to theater attendance; predicting the future of the English curriculum; the composing processes of twelfth graders.

It is obvious from this summary that the research is widely diverse, and like the subject of English itself, "research in the teaching of English" is hard to define. Nonetheless, the research has tended to concentrate on traditional concerns of the English curriculum—the teaching of language, literature, and composition. There is a paucity of actual research (though much writing) on widely discussed recent concerns and developments such as effects

[5] K. W. Hunt, *Grammatical Structures Written at Three Grade Levels* (Champaign, Ill.: NCTE, 1965) [Originally published as CRP #1998, USOE]; R. G. Strickland, *The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children* (Bloomington, Ind.: School of Education, Indiana University, 1962); W. Loban, *The Language of Elementary School Children* (Champaign, Ill.: NCTE, 1963) [The study was originally published in a report to the USOE in 1961]; C. Chomsky, *The Acquisition of Syntax in Children from 5 to 10* (Cambridge, Mass.: M.I.T. Press, 1969).

[6] W. Labov, *The Social Stratification of English in New York City* (Washington, D.C.: Center for Applied Linguistics: 1966).

of new curriculum structures on learning in English, use of dramatic activities, use of technology, and effects of and response to non-print media. Certainly research in language development, particularly syntactic control, represents a heavy emphasis, and research in this field, coupled with study of applications to teaching of linguistics, dominates the research that has surfaced for special recognition.

Emphases in Design

What of current trends or emphases in research design or genres of research? Of the eleven studies reported in the 1971 issues of *RTE,* six were experimental, two descriptive surveys, two relational or correlational studies, and one a cross-sectional analysis. The experimental study, long a staple of educational research, is far less prominent in the research that has received special recognition through the NCTE. None of the distinguished research awards for specific studies has been given for an experimental study. Four have gone to basically cross-sectional analyses of language development, and one for a work of literary criticism. Only one of the twelve promising researcher citations has recognized an experimental study. Six have been given to cross-sectional analyses of performance, four to relational studies, and one to a content analysis. Three of the NCTE research monographs are based on experimental studies, two on a longitudinal analysis (Loban), five on cross-sectional analyses, and one each on historical survey, descriptive survey using a questionnaire technique, and case studies.

The prominence of cross-sectional analyses in recognized research correlates with the recent emphasis on language acquisition and development. The research of Kellogg Hunt, Ruth Strickland, and Roy O'Donnell, William Griffin and Raymond Norris[7] has set a general pattern for many other studies. The relative absence of experimental studies in recognized research, despite the prominence of the genre generally, underscores the fact that good experimental designs are simply harder to build than other types. The experimental study—usually involving the

[7] R. C. O'Donnell, W. J. Griffin and R. C. Norris, *Syntax of Kindergarten and Elementary School Children: A Transformational Analysis* (Champaign, Ill.: NCTE, 1967).

effects of teaching methods, curricular arrangements, or materials—have been dogged with especial problems which may or may not be insurmountable in research in the teaching of English. Longitudinal studies are rare for the obvious reason that they require vast expenditures of time and resources. Funding for studies such as that of Loban is hard to get, and many researchers, especially doctoral students and young professors hoping to earn promotion, are loath to invest eight or ten years in any one research. Case study technique has been out of fashion generally in most educational research for some years—a fact which this writer bemoans later in this paper.

General Criticisms and Problems of Research

Though there are some evidences of cumulative progression in recent research related to the teaching of English—that in certain aspects of language development, for example, and the sequence of important studies by Bateman and Zidonis, Mellon, and O'Hare [8] on the effects of sentence-combining techniques derived from transformational grammar on writing performance—many critics have pointed to the fragmentary, isolated, unrelated nature of much of the research, a problem shared no doubt by other fields in education. The basic problem is one of dissemination. Most doctoral research, still a significant part of the total research in the field, does not get beyond the supervisory committee, except for an occasional inter-library loan, and many of the projects, particularly in the small grants category, sponsored by the U.S. Office of Education end in undisseminated reports to Washington. The NCTE publication program, including *RTE,* has helped much, but the major breakthrough was represented by the establishment of the NCTE/ERIC which recently has been combined with the former clearing house on reading based at Indiana University and is now titled Clearinghouse on

[8] D. Bateman and F. Zidonis, *The Effect of a Study of Transformational Grammar on the Writing of Ninth and Tenth Graders,* (Champaign, paign, Ill.: NCTE, 1966); J. C. Mellon, *Transformational Sentence Combining* (Champaign, Ill.: NCTE, 1969); F. O'Hare, The effect of sentence combining not dependent on formal knowledge of a grammar in the writing of seventh graders. Unpublished Ph.D. dissertation, Florida State University, 1971.

Reading and Communication Skills. The impact of the clearing house operation is just beginning to be felt.

The participants in a research seminar held at New York University in 1963 identified some problems that are still much in evidence.[9] A major one is that change in the significant competencies associated with achievement in English come slowly. Short-term experimentation of a few weeks or months is often doomed to conclude with the tired refrain, "no significant differences." The participants pointed out, too, that since so much research has been dismissed as trivial or inconclusive, researchers frequently attempt to take on too much in one study in order to obtain more widely generalizable results. Careful generalization in terms of very specific poulations need not vitiate studies, but a time-honored fault in research is to generalize beyond the specific data gathered.

Most of the criticism of research design has centered on the experimental study, an indication both of the high status and inherent problems of experimental research. Doris Gunderson has summarized succinctly some basic faults in design of experimental studies in English.[10] She points particularly to lack of specificity in basic postulates or assumptions and to the frequent lack of explicit, logically derived hypotheses. She cites, too, the familiar problems of sampling—incompleteness, lack of a specific sampling plan, and lack of well-defined populations. Control of variables is, of course, the eternal ghost haunting experimental studies. In most educational studies, complete control of variables that could affect outcomes is impossible, though many designs could be tighter than they are. It is, in fact, the problem of control of variables that makes many researchers settle for designs that can only produce trivial results.

Walter Petty cites a serious, related problem of experimental studies of methods or teaching strategies: "A major flaw in most of the studies examined . . . was the lack of specificity as to the part a particular method or procedure played in a study. The studies simply did not satisfac-

[9] L. M. Rosenblatt, *Research Development Seminar in the Teaching of English* (New York Univ.: 1963). [A report to the USOE.]

[10] D. V. Gunderson, "Flaws in research design," *Research in the Teaching of English*, (Spring, 1967), pp. 10–16.

torily compare methods . . . possibly such comparisons shall continue to be absent from the research scene, not because methods could not be detailed specifically, but because the skills needed for learning from the application of different methods may well be so different, so lacking in similarity, even, that they are not comparable. These skills, too, may be learned and held with varying degrees of success by different individuals due to many, as yet unexplored, factors."[11]

Naivete in use of statistical procedures and inappropriate statistical techniques long have been frequently mentioned shortcomings of research in the teaching of English. Most researchers today have statistical expertise and computer assistance in analysis of data at hand. Even so, some students of research design maintain that many studies, otherwise well designed, use statistical analyses that have insufficient power to detect significant effects or changes in terms of the magnitude of the effect and the number of subjects. "No significant difference," then, may be sometimes purely a function of the analysis used and the interpretation of its results.[12]

What Follows

Following this general overview, this paper will attempt to identify in the next section some of the more specific accomplishments and problems in several arbitrarily selected substantive areas of research in the teaching of English: 1. Curriculum; 2. Language development; 3. The teaching of written composition; 4. The teaching of literature; 5. The teaching of oral language and use of dramatic activities. No attempt is made to review the research in each of these areas in a comprehensive sense, nor is there any claim that the "best" studies necessarily have been identified. "Research," for the purposes of this paper, is defined generally as that fitting into the categories identified above—status surveys, relational and correlational studies, longitudinal studies, cross-sectional analyses, experimental studies, and case studies.

The final section will suggest some possibilities and

[11] Petty, op. cit., pp. 84–85.

[12] Based on discussion with Prof. F. J. King, Florida State University.

make some recommendations for future research in the teaching of English.

RESEARCH ON CURRICULUM DESIGN AND STATUS

The rather awkward double title of this subsection—"design and status"—suggests its organization: first, a statement of what has not been done and then a statement about some useful research on curriculum status. The lack of impact of research on curriculum design has become a refrain in this paper as it is in other general statements about research in the teaching of English. Despite much recent experimentation with curriculum patterns in English in the schools, there has been virtually no research on the connection between design and learning outcomes.

The federally supported curriculum study centers of the 1960's might have helped to fill this gap, but they did not. Most of them were materials producing enterprises (valuable, of course, in that function), and though there were a few attempts to devise new instruments to measure what was taught in particular programs, there was no systematic effort to evaluate the success of most of the programs or to compare their effects with those of curriculum designs they were to supersede. Furthermore, the theory upon which most of the programs were based, the Brunerian concept of cumulative sequence, was being questioned widely even before many of the centers had finished their work.

Perhaps the most widespread and fashionable innovation in curriculum design in secondary school English in recent years is that of short-term electives, in which traditional grade levels may or may not be preserved, a design given much publicity through the publications from the Trenton, Michigan, High School on its "non-graded phase elective curriculum" [13] Yet again, evidence on the success of such programs, especially in terms of student learning, is scarce. In his survey of elective programs in English, published in 1972, George Hillocks, Jr. found that few of the programs featured systematic evaluation (though he noted that most traditional programs probably make even "feebler" efforts at evaluation).[14] Department

[13] *Project APEX* (Trenton, Michigan Public Schools, 1970).

[14] G. Hillocks, Jr. *Alternatives in English: A Critical Appraisal of Elective Programs,* (Urbana, Ill.: ERIC/RCS, 1972).

chairmen and supervisors connected with elective programs nearly unanimously reported in the Hillocks survey that teacher and student attitudes were heightened in elective programs, and a few reported such benefits as higher student grades, fewer discipline problems, and greater student enrollment in English courses. Hillocks concluded that on the basis of the scant evidence available, elective programs provide no advantage over traditional programs in terms of student growth in measurable cognitive areas, and he noted that many of the courses in programs included in his survey parallel college and traditional high school offerings and appear to reflect teacher rather than student interests.

Curriculum designs such as the short-term elective, along with important curriculum theory like that of James Moffett who labels his work "a chart for further exploration and a kind of rallying call," [15] need testing through experimental designs so that schools may really know what they are rallying to. It seems strange, on the one hand, that more schools, especially college and university laboratory schools, have not been active in school-wide research on curriculum design. Yet, on the other hand, some explanations seems obvious. A first lies in the old problem of agreeing upon significant outcomes and of valid means for measuring or assessing them. The behavioral objectives movement may or may not have alleviated these problems—it seems too early to tell. A second explanation is that teachers and administrators are not generally research oriented, and administrators are loath to risk large groups of students in long-term experimentation. Sheer logistics furnish a further explanation. It is not easy, financially and otherwise, to assemble the kinds of research teams needed to carry out such experimentation. The curriculum center projects presented the golden opportunity to solve this last problem, but perhaps another golden age of federal funding for research in the teaching of English may come.

Status Surveys

The most valuable research on general curriculum in English probably is represented by the status surveys, the most widely known of which are the "national interest"

[15] J. Moffett, *A Student-Centered Language Arts Curriculum, Grades K-13: A Handbook for Teachers,* (Boston: Houghton Mifflin, 1968).

reports of the NCTE[16] and the Squire-Applebee studies of selected secondary English programs in the United States and in England.[17] The national interest reports were political documents, designed to stir action on the federal front, rather than research publications, but they did draw on status research and presented information on a national basis highly valuable at the time. Certainly they were a major force in generating the English program of the U. S. Office of Education.

The federally supported Squire-Applebee studies of English curriculum and teaching methods in a broad sampling of high schools judged to be superior furnished important information indicating directions for improvement in English programs and in teaching conditions. The findngs, frequently disheartening to many English educators, were given greater validity through the methods used in gathering data, involving site visits and classroom observations rather than mere summary of information gleaned from responses to questionnaires.

Curriculum improvement is dependent, of course, on continuously current and accurate information on the status of English teaching on a national scale. John C. Maxwell, Associate Executive Secretary of the National Council of Teachers of English, has suggested a possible program which could help to assure a continuous flow of such information. He suggests that NCTE agencies, working with colleges and universities, encourage and give some support and publicity to status studies by graduate students in various regions of the country.[18] The design of the studies would be the responsibility of local graduate advisors.

Edmund Farrell's unique dissertation predicting the future might be considered a status study of the future English curriculum, 1970–2000,[19] and may offer as reliable

[16] *The National Interest and the Teaching of English,* (Champaign, Ill.: NCTE, 1961); *The National Interest and the Continuing Education of Teachers of English,* (Champaign, Ill.: NCTE, 1964).

[17] J. R. Squire and R. K. Applebee, *High School English Instruction Today,* (New York: *Appleton-Century-Crofts,* 1968); J. R. Squire and R. K. Applebee, *A Study of the Teaching of English in Selected British Secondary Schools.* Final Report, USOE Project.

[18] The very tentative suggestion was made in a letter to the author and several others in the Spring of 1972.

[19] E. J. Farrell, *Deciding The Future: A Forecast of Responsibilities of Secondary Teachers of English,* 1970–2000 AD (Urbana, Ill.: NCTE, 1971).

a guide as can be charted, for planning for the future. The forecasts of his panels of specialists are not particularly startling, representing an extension of present trends. The non-English specialists—in learning theory, educational technology, and curriculum—foresee more accurate and significant testing, extensive use of computer-assisted and other electronically modulated instruction, modular scheduling, differentiated staffing, and multi-level ungraded programs. (In addition, they think most classroom teachers will not be prepared to carry out "action" research related to their teaching problems by the end of this century!) The English specialists foresee more flexible types of curriculum with clearly defined behavioral objectives, greater variety in the content and organization of literature programs, broadened language study, greater attention to processes underlying oral and written composition, and greater emphasis on multi-media learning. Farrell's use of the "Delphi technique," in which progressive questionnaires evolve from the response of his panel, is worthy of study by researchers who plan to use questionnaires.

RESEARCH ON LANGUAGE DEVELOPMENT

One can only be impressed, and a bit overwhelmed, by the bulk of research in language development—only one of the areas of psycho-linguistics research—if only after a reading of such reviews as those by Griffin and MacGinitie.[20] As indicated in the previous section, research in syntactic development has dominated that selected for special recognition by the NCTE. Only a few comments are possible here on the substance and method of research that has as yet important, but largely unrealized, potential for the teaching of English.

Despite the quantity of research in the field, it is interesting that comparatively little has been done in the area of development in transformational-generative semantics ("denotative" as contrasted to "connotative" or Hayakawa semantics—a distinction borrowed from Kellogg W. Hunt), probably because transformational semantic theory is a later

[20] W. J. Griffin, "Children's development of syntactic control," in *Developments in Applied Psycho Linguistics Research*, ed. by Sheldon Rosenberg and James H. Koplin (New York: Macmillan, 1968), pp. 19–65; W. M. MacGinitie, "Language development," in *Encyclopedia of Educational Research*, Fourth Edition (New York: Macmillan, 1969), pp. 686–699.

development than transformational syntactic theory, and the sophistication of study of language development has had to proceed hand in hand with the development of systems for analyzing the syntactic and semantic components of language. The recent research of Jeremy Anglin, however, giving a limited profile of the development of the lexicon from grade three to adulthood, represents important advances in study of semantic development,[21] and other studies are in progress such as a doctoral dissertation at the Florida State University on the progression in acquisition of noun features.

MacGinitie's classification of three basic methods are "observation," "testing," and "teaching." Using observation are case studies and cross-sectional analyses of samples of oral or written languages gathered from natural or unstructured situations in which language is used. Testing is the central method in those studies which structure a particular situation in which language must be used in order to discover, for example, whether children at given ages can use particular syntactic constructions or know the meaning of words. A third group of experimental studies assesses the effects of attempts to teach or condition language behavior in some way. Studies in this latter category encounter, of course, the usual problems of sampling and control of variables, and the variables affecting language development may be even more diverse and harder to control than those affecting other outcomes in the teaching of English.

Much of the research in language development has concentrated on young children; comparatively little has been done with adolescents. Carol Chomsky's recent research, though restricted to the age span from five to ten, shows that there is a surprisingly late acquisition of some syntactic structures and suggests that active syntactic acquisition may take place well beyond the age of ten.[22]

MacGinitie cites a shortcoming of many of the studies—the use as subjects of unusually bright children who may telescope some stages of language development or pass through them so quickly that they are not noted by the observer.

[21] J. M. Anglin, *The Growth of Word Meaning*, (Cambridge, Mass.: M.I.T. Press, 1970).

[22] Chomsky, op. cit.

Implications for teaching and the curriculum are the major interest of English educators, of course, in research in language development. Kellogg W. Hunt has suggested a basic objective for applied research—the "hastening and assuring" of syntactic maturity in language use (there may be some difference of opinion on the desirability of "hastening")—and suggests that a basic question for research may be "At what age does curriculum X or teaching method Y have a beneficial effect on the use of language?" [23] Hunt maintains that there are certain sentence transformations which children more or less automatically learn as a function of increasing age. Hunt would have researchers discover what these transformations are so that they could be taught at an earlier age than that at which they seem to be used naturally. His research shows, for example, that older students use the "agent deletion" transform and write "Aluminum is extracted from bauxite" instead of "Workmen extract aluminum from bauxite" (making the decision that it is not important to indicate that workmen do the extractions).[24]

Of course, if syntactic maturity is to be "hastened and assured," there is an assumption that a valid index of syntactic maturity is available—and there is: Hunt's "T-unit" index which shows a significant progression from fourth graders to superior adult writers in the length of terminable units and the number of reduced sentences embedded within a main clause.

Some important research already has been carried out in the vein that Hunt suggests, notably the experimental studies of John Mellon and Frank O'Hare.[25] Both researchers found that syntactic maturity in the writing of seventh graders was significantly affected by instruction in sentence embedding.

Herbert Karl suggests that Hunt's index should be supplemented by some assessment of ability to use language as an "explicit vehicle for logical thought." In the conclusion

[23] This material is abstracted from conversations of the author with Professor Hunt.

[24] K. W. Hunt, "Teaching syntactic maturity," in *Applications of Linguistics, Selected Papers of the Second International Congress of Applied Linguistics,* (Cambridge, England: Cambridge Univ. Press, 1969), pp. 287–300.

[25] Mellon, *op. cit.*; O'Hare, *op. cit.*

of his dissertation on curricular implications of Jean Piaget's work, Karl says: "If, as Piaget seems to suggest, there is a gradual development of the child's ability to understand and use language for explicitly logical purposes, then it would be useful to learn the extent to which such logical maturity parallels syntactic maturity." In order to find this out, Karl proposed that instruments be constructed to measure the progression in logical maturity on the basis of indices posited by Piaget: (1) the "because" of casual explanation; (2) the "because" and "since" of logical justification; (3) the "although" of discordance; (4) the "if . . . then" of implication; (5) the "either. . .or" of disjunction. Karl suggests that correlational studies of the relationship of syntactic and logical maturity would be illuminating.[26]

Important implications for teaching and curriculum have been uncovered, too, by research such as that by Loban and Labov which shows the relationship of language development and problems to race and social class.[27] The teaching of standard English as a second dialect can draw upon a considerable body of research on the non-standard features of the speech of blacks and whites and of various ethnic groups. Loban's longitudinal study, the only one of its kind, stands as a landmark for other needed long-term studies.

RESEARCH IN THE TEACHING OF WRITTEN COMPOSITION

Rhetorician Walker Gibson wrote in 1969 that "our teaching of English composition, by and large, has for years been a shambles. In the schools it is the area where teachers feel most at sea, confessing themselves most in need of self-confidence and assistance. In the colleges, especially in the universities, it falls characteristically into the least experienced hands, where it is pawed and plied into a thousand inchoate shapes. The composition teacher, as everyone knows, can show no respectable theory; his discipline boasts no scholarship but is planned by dolts, manned by drudges, and avoided if possible by everyone." [28] Even though things may not be *that* bad, Dean

[26] H. Karl, The development of language in children: an analysis of selected works of Jean Piaget with implications for the English program, K-9. Unpublished Ph.D. dissertation, Florida State Univ., 1971.

[27] Loban, *op. cit.*; Labov, *op. cit.*

[28] W. Gibson, "Composition as the Center for an Intellectual Life," in *Hues of English, NCTE 1969 Distinguished Lectures* (Champaign, Ill.: NCTE, 1969), p. 75.

Memering, in a recent dissertation, was able to find little research support for recent theories and approaches in the teaching of composition, though he found that teaching approaches could be classified according to four general rationales.[29]

In view of pronouncements and findings such as these, it is not surprising that reviews of research in the teaching of composition generally have echoed an especially doleful tone, centering on needed research and the problems in the field rather than on accomplishments, though the landmark publication in 1963 of *Research in Written Composition* by Braddock, Lloyd-Jones, and Schoer could review in detail five exemplary studies and point to some profitable lines of investigation.[30]

Only relatively minor updating, actually, of the Braddock report seems appropriate as Braddock's section in the *Encylopedia of Educational Research* indicates.[31] The report was pessimistic about the relation of research done at that time to the big central question which the editors posed: "What kinds of writing following what kinds of instruction for what kinds of students?" The report stakes out a considerable expanse of "unexplored territory," represented by twenty-four questions, the last of which is "Of what does skill in writing really consist?" This basic question, and others more specific, suggested to the editors the usefulness of case studies, and in the report some evidence is summarized on the differences in approaches in composing of writers adjudged to be good or poor.

The report deals with the problems of assessing performance in composition including the troublesome variables of the writer himself at a given time when he writes, of particular assignments in writing, and of the rater or evaluator. The report calls for improvements and greater imagination in "frequency count" research on student writing. A failing, common to much experimental research, particularly that of doctoral students, is noted: lack of trial runs or pilot studies which might eliminate the "bugs" from designs and instruments for evaluation.

---

[29] D. Memering, *Recent theories and practices in the teaching of composition*. Unpublished Ph.D. dissertation, Florida State University, 1971.

[30] Braddock, Lloyd-Jones, and Schoer, *op. cit.*

[31] R. Braddock, "English composition," in *Encyclopedia of Educational Research*, Fourth Edition (New York: Macmillan, 1969), pp. 443–461.

Few hard conclusions resulted from the review of research on environmental and instructional factors influencing performance in composition: improved performance is generally related to amount of practice; revision may help; study of grammar does not help. On the latter point, though, the participants at the seminar on research in the teaching of English held at New York University in the same year the Braddock report was published asserted that many studies failed to show the relationship between study of grammar and improved composition because they did not measure "truly relevant aspects of either grammatical knowledge or composition performance."[32] The participants did not spell out ways of righting the wrong, though they wondered wistfully how relatively independent skills and abilities, or "aspects or proficiency" in English, might be identified, possibly through "factor analytic methods."

The relationship of personality factors to the composing process and to performance in writing (are there innate personality traits that make better writers of some people than others?) considerably interested Henry Meckel in his section in the *Handbook of Research in Teaching*, though the research he cites is sparse.[33] There seems to have been little concern with this problem recently. For example, it is not reflected in William West's review in 1967.[34] West cites twenty-one references under these categories: teaching methods; stimuli for writing; teachers' attitudes toward teaching composition; and the relation of grammar study to performance in composition. Bemoaning the backwardness of research in the field, West shrugs off research in methods of teaching composition which relies on evaluation of student writing since he thinks that valid means of evaluating writing or of assessing growth in writing had not been developed as of 1967.

Though grading and evaluation of student compositions has been a principal use of English teachers' time for many years, it is obvious that a major bugaboo of research in the

[32] Rosenblatt, *op. cit.*, p. 33.

[33] H. Meckel, "Research on teaching composition and literature," in *Handbook of Research on Teaching*, ed. by N. L. Gage (Chicago: Rand McNally, 1963), pp. 966–1006.

[34] W. West, "Research on written composition," *Review of Educational Research*, XXXVII (April 1967), pp. 159–167.

teaching of composition is lack of agreement on valid methods of assessing performance and growth in writing. A second is lack of solid knowledge of the nature of the composing process.

In their publication of 1966, Godshalk, Swineford, and Coffman report a study which, in the view of the College Entrance Examination Board, was a "breakthrough" in measurement of writing ability. The major conclusion of the study is that "The most efficient predictor of a reliable direct measure of writing ability in one which includes essay questions or interlinear exercises in combination with objective question." [35] (The interlinear exercises require students to correct faults in running text.) The study was promptly attacked by Martin Steinmann as conceptually unsound because based on an inadequate conception of writing ability.[36]

In a surface sense, at least, the problem of evaluating student writing may have been overemphasized. "What is good writing?" is one of those questions akin to "What is the good life?" But if, for a given experiment, the criteria for quality or improvement in writing are agreed upon by those who will rate the students' writing, there seems little doubt that a fairly high reliability of multiple ratings can be obtained if certain procedures are followed.[37] A very real problem, though, is measurement of small increments of growth in writing ability. Many studies which have featured experimental treatment over a period of only a few weeks or months have been predestined to conclusions of "no significant differences," since one thing that is known, at least, is that improvement in general aspects of writing ability is a slow, gradual process. Attacking the problem takes us back, of course, to criteria for evaluating writing, to factors influencing writing performance, and to the nature of the composing process, but it may be possible to identify those aspects of writing ability that can be expected

[35] F. I. Godshalk, F. Swineford, and W. E. Coffman, *The Measurement of Writing Ability*, (New York: College Entrance Examination Board, 1966), p. 41.

[36] M. Steinmann, "A conceptual review," *"Research in The Teaching of English*, (Spring 1967), pp. 79–84.

[37] P. B. Diederich, "How to measure growth in writing ability," *English Journal*, 55 (April 1966), pp. 435–449.

to show growth in short periods. Sentence embedding may be one example.

Perhaps the most significant recent study of the composing process is that of Janet Emig.[38] Responding to a challenge in the Braddock report and to other calls for needed research, Emig's research, involving case studies of eight twelfth graders, investigates the ways that students usually or typically behave as they write. Emig analyzed the comments recorded by the eight students as they composed three short themes aloud. Further, the students produced "writing autobiographies," composed of their recollections of earlier writing experiences that were rewarding or painful for some reason. The research reveals, too, some of the differences that school-sponsored or teacher-assigned writing (as opposed to self-sponsored writing) made on choice of topic, selection of material, degree of personal engagement, and attention to planning, pre-writing, and revision.

Emig's case-study techniques are worthy of study for their possible use for a variety of other purposes. For example, Earl Buxton suggests, in his introduction to Emig's monograph, that Emig's methods might be used to study student responses to different kinds of writing instruction, a kind of study that might be more illuminating than that of study of methods in an experimental design (though the two might well be combined in some studies).

Emig's study raises again the complicated question of the relationship between composing orally and in writing. How can the student be helped to frame effectively in writing the thoughts he composes orally? Such approaches as Robert Zoellner's much debated "talk-write" pedagogy[39] need further investigation.

RESEARCH ON THE TEACHING OF LITERATURE

This section is based largely on the Purves-Beach report[40] which was made available to the writer in manuscript form while this paper was being prepared and which was published in late 1972.

The report, in general, is more optimistic about the

[38] J. Emig, *The Composing Processes of Twelfth Graders,* (Urbana, Ill.: NCTE, 1971).

[39] R. Zoellner, "Talk-write: a behavioral pedagogy for composition," *College English,* 30 (January 1969), pp. 267–320.

[40] Purves and Beach, *op. cit.*

quality and accomplishments of the research in the teaching of literature than is the Braddock report on written composition, though many limitations of what has been done and many gaps that might be closed by research are cited. The emphasis in the report is on the substantive aspects of research rather than on technique, but there are many incidental references to design and technique. In that respect, the authors seems to be more interested in designs other than the experimental—perhaps reflecting a general trend.

The report is organized in three divisions: studies of response to literature, studies of reading interests, and studies of the teaching of literature. The amount of research on response to literature surprised the authors as it will readers of the report, probably accounted for by the fact that many of the studies are doctoral dissertations that have not been widely disseminated. The report verifies a common impression that studies by Squire and by Purves and Rippere [41] validly have established the categories of responses which students express orally or in writing to literature they read, and many studies have been based on these categories.

Although finding fairly strong evidence that "the interest and preoccupations of the reader are important ingredients in literary response, preference, and understanding," the report goes on to say that "beyond this general conclusion we still know little about the relationship between interest and understanding, personality and understanding or response, and environment and response." Though the general forms of response may have been established, as noted above, the report cites very few studies on the *process* of an individual's response, "what happens to the reader from when he picks up the work to when he finishes it," and the authors suggest case studies to find out more about this process. One might suggest here, too, that case studies might furnish needed information on immediate as opposed to deferred responses. Most studies have been limited to more or less immediate response.

[41] J. R. Squire, *The Responses of Adolescents While Reading Four Short Stories,* (Champaign, Ill.: NCTE, 1964); A. C. Purves and V. Rippere, *Elements of Writing about a Literary Work: A Study of Response to Literature,* (Champaign, Ill.: NCTE, 1968).

Disappointing to many theorists of the literary experience will be the report's conclusion that there is no clear evidence that people change behavior as a result of reading literature. People claim to, but do they, the report asks. There is evidence, however, that clearly neurotic people gain therapy from literature, but people who are not neurotic are a different matter. Case studies again are suggested as the best way to find out whether literature is broadly therapeutic.

The report points to the need for research on the relationships of various aspects of response, between catharsis and emotional pleasure, for example, and suggests that such studies "might well combine the case-study technique of exploring many aspects of the responses of a few individuals . . . with multi-variate analysis, multi-dimensional scaling, partition analysis, and other, more sophisticated statistical treatments."

The report tends to substantiate a point made earlier in this paper—the paucity of research on response to and rewards from literature in nonprint media as compared to those from reading literature. One dissertation, just completed at this writing, was a comparison of responses to narrative and lyric short stories and films, finding, for example, significantly more interpretational responses to film and narrational responses to short stories.[42]

One rather clear implication of the report is that, if research on response is to be useful in determining sequence in literature study, a large-scale longitudinal study would be of great value, possibly providing answers to such questions as: at what point does a significant proportion of students respond to an element such as irony? To some educators, sequence in literature study is properly a myth anyway, but if there is to be improvement on the present rudderless sequences, directions may lie in knowledge of developmental patterns of response.

The report reflects the obvious volume of research on reading interests, though many of the studies are small-scale and based on sampling that makes impossible any broad generalizations from findings. Most studies deal

[42] W. J. Lewis, A comparison of student responses to lyric and narrative films and short stories. Unpublished Ph.D. dissertation, Florida State University, 1972.

with preferences among titles or categories of content—boys like war stories, girls like mysteries at certain age levels, for instance. Little has been done on rewards of literature to the reader at different stages. The authors attribute this lack mainly to the absence of content analysis schemes which could be applied profitably to the "fad" books that sweep through schools each year. What did readers really get from *Catcher in the Rye, One Flew Over the Cuckoo's Nest,* or *Airport* that put the books in the fad category at a particular time? The report cites the need for clarification of the relationship between interest and difficulty: "does high interest result in ignoring difficulty or does immediately perceived difficulty deter development of interest in the book?"

Purves and his colleagues noted with surprise the relatively meager research on the teaching of literature. The report classifies the studies, most of them experimental, in six categories: (1) those in which the treatment variable is material; (2) those in which the treatment variable is an instructional technique; (3) those in which literature is the treatment variable for some instructional end such as reading or writing; (4) evaluations of curriculum; (5) studies of extensive reading; (6) studies of teachers and teacher education.

Evaluation of teaching methods or curriculum in literature in terms of student behavior or achievement is as thorny a problem as it is in written composition. The report lauds the work in evaluation of the Florida State University Curriculum Study Center, the Carnegie-Mellon Curriculum Study Center, and the Fader-McNeil project,[43] concluding that "If one were to pool their measures, one would have a substantial and effective criterion measure for curriculum evaluation. It is to be hoped that future projects make use of these measures."

The section on research in the teaching of literature concludes with recommendations for further research in two areas: (1) response patterns and achievements that

[43] *The Development and Testing of Approaches to Teaching of English in the Junior High School.* Report to USOE on Project No. H-026; Final Report of the USOE on Project No. H-014; D. N. Fader and E. McNeil, *Hooked on Books: Program and Proof,* (New York: Berkeley, 1968).

emerge from study of ethnic literature on the part of students from the ethnic group represented by the literature and from other ethnic groups; (2) the "incidental effects" of literature teaching—on linguistic development or on response to other arts, for example.

RESEARCH ON ORAL LANGUAGE AND DRAMATIC ACTIVITIES

The primacy of oral language in communication and its involved relationship with other aspects of language development have long been recognized, but attention to the research that has been done and should be done in the field may have been given added impetus by recent and urgent calls for greater emphasis on oral language and dramatic activities in the classroom. Since there seem to be no problems of research design that are not relatively common to other aspects of the teaching of English, this section will be principally a brief commentary on some substantive accomplishment and possibilities of research in the oral and dramatic area.

A committee of the National Conference on Research in English, in a publication in 1967, subsumes research in oral language under these categories: (1) Oral Language and Personal and Social Development (2) Oral Language and the Development of Other Language Skills (3) Effects on Environment on Oral Language Development (4) Listening (5) Evaluation of Oral Language Performance.[44] The report emphasizes the fact that oral language serves as the base for the development of achievement in reading and writing, and two of the authors conclude that "Without a full understanding of listening and speaking, the two basic skills in language can never become any better than they exist today."[45] And the editor of the publication, Walter Petty, asks whether we really know which phonological, syntactical, or lexical deviations interfere with communication through their effects on the listener.[46] Such caveats apparently have stirred some research action since a number of recent studies, including some of those recognized

[44] W. T. Petty, editor. *Research in Oral Language,* (Champaign, Ill.: NCTE, 1967).

[45] H. E. Blake and A. J. Amato, "Needed research in oral language, in *Research in Oral Language,* p. 65.

[46] W. T. Petty, "Needed research in oral language," in *Research in Oral Language,* p. 66.

by the NCTE promising researcher citations, have dealt with interrelationships of major language skills.

As indicated in an earlier section, research has rather clearly identified the problems with standard English and the non-standard markers in the speech of children from homes in which non-standard English is habitually spoken. Loban's study, for example, shows that the principal problems are with irregular verbs and that ten years of schooling brings no standardization in use of verb forms among such children. Such evidence lays a basis for training children to use an alternate standard dialect, but "bidialectism" has become philosophically controversial. Ruth Strickland foreshadowed the present controversy when she said, in the NCRE publication cited above: "Which regional variations to correct or modify and which to accept as part of the seasoning that marks American English as interesting, vital and colorful is something researchers could well study."[47]

The vigorous research activity in listening is a curious phenomenon. It has run its course for several decades in a virtual vacuum despite periodic lamentations of the neglect of listening in the classroom. Status studies continually have shown practically no direct attention to listening in the English curriculum, though the NCTE insisted for years as a policy line that the components of the curriculum were reading, writing, speaking, and listening.

In a recent NCTE/ERIC report, Sara Lundsteen synthesizes a considerable amount of the research in listening, illustrates the application of some of it in teaching, and finds a need for further study in these areas: (1) the personality adjustment dimension of listening (2) utilization of compressed (and expanded) speech (3) growth patterns (4) skills—how to teach them and with what materials (5) listening vocabulary.[48]

Regarding technique, she suggests: "Existing studies might be replicated and future studies devised with more careful criteria, such as reports in detail of how the teacher

[47] R. G. Strickland, "Needed research in oral language," in *Research in Oral Language*, p. 61.

[48] S. W. Lundsteen, *Listening: Its Impact on Reading and the Other Language Arts.* (Urbana, Ill.: NCTE/ERIC, 1971).

is behaving, how the pupil is actually behaving, and what the interaction is. Evidence to verify these behaviors might be, for example, from video tape, observation scales administered by trained personnel including both verbal and nonverbal behaviors, or stenographic records. Collection of such evidence and placement in a data bank would allow investigators to use the data for purposes other than those originally intended."[49]

Lavish claims by the proponents with little or no empirical substantiation necessarily must be the scholarly summary to date of the value of dramatic activities in the *English curriculum*, despite the furor over drama *in* the classroom, though the furor may be more in print than in practice, especially on the American side of the Atlantic. Exemplifying the evangelism of proponents is Joseph Karioth who, in *Elementary English* for February 1972, identifies "certain behavior patterns that can be expected as a result of dramatic experience." Among these behaviors, according to Karioth, are expanded "sense awareness"; successful expression behavior"; articulation and inflectional skills, vocabulary building, and language acquisition"; "an understanding and knowledge of the kinds of patterns of behavior that are expected in particular role relationships"; "creative behavior that define one's individuality."[50]

James Hoetker, in an NCTE/ERIC monograph, discusses some of the major types of dramatic activities that may be viable in the classroom and reviews the sparse research related to each.[51] He is able to cite two experimental studies which found that groups of students given extensive experience in creative dramatics developed greater ability in creative thinking than did the control groups. Speaking of the extravagant claims for the benefits of dramatic activities, Hoetker says: "The idea that drama can contribute to development in these areas is, besides being supported by the testimony of experience, commonsensical.

---

[49] *Ibid.*, p. 100.

[50] J. Karioth, "Drama and the elementary school child," *Elementary English*, XLIX (February 1972), pp. 302–304.

[51] J. Hoetker, *Dramatics and the Teaching of Literature*, (Champaign, Ill.: NCTE/ERIC, 1969).

But—and this always needs to be emphasized—the authorities agree that development through drama is a gradual cumulative process, and it is very uncertain what may be the developmental timetable in each area, especially if drama is only an occasional activity." [52] Hoetker's point suggests the potential value of a longitudinal study of a group of pupils who are involved in a variety of dramatic activities over a period of several school years.

Some of the enthusiasts for dramatic activities are interested specifically in the contributions to the study of literature. Dramatic improvisation, especially, has been emphasized as an important kind of readiness for reading imaginative literature. The English educator, J.W. Patrick Creber, for example, writes: "It must be clear that many novels offer situations that lend themselves to dramatic treatment, but it is worth noting that this is essentially a two-way process, for the acting out of such situations may be expected to modify and deepen the actors' comprehension of the novels from whence they were taken. Furthermore the children's receptivity to a particular play may sometimes be notably improved by dramatic improvisations on some of the situations, themes, and characters it contains, before the play itself is read."[53]

It should be quite possible, through experimental studies or case studies, to test hypotheses such as those posed in this statement.

SOME RECOMMENDATIONS FOR THE IMMEDIATE FUTURE

This part of the paper is easy to write, of course, being bounded only by the limits of the writer's imagination and common sense. Lists of needed research abound, and another would be superfluous here, though a rather formidable list could be compiled on the basis of the content of the preceding parts of this paper. What follows here merely pinpoints a few matters which the writer thinks are especially worthy of immediate attention on the research front.

---

[52] *Ibid.*, p. 29.

[53] J. W. P. Creber, *Sense and Sensitivity,* (Univ. of London Press, 1965), p. 63.

Status Surveys

TWO SUGGESTIONS ON LOGISTICS OF RESEARCH

Reference was made in the preceding section to the suggestion by John Maxwell, Associate Executive Secretary of the NCTE, that the NCTE encourage, coordinate, possibly subsidize, and disseminate findings from a continuous program of status surveys of the teaching of English by graduate students in different regions of the country. Such a program, which would be most useful, could be carried on quite feasibly by the Committee on Research and the Research Foundation of the NCTE through liaison with persons who direct graduate student research at various institutions.

Research Centers

Similarly, various universities or colleges, or consortiums of colleges or universities and public school systems, might be designated as centers of research in certain aspects of the teaching of English. Again, the NCTE, working with the universities or consortium through the Committee on Research or the Research Foundation, might identify such centers. Together, the centers and the NCTE might solicit federal or private foundation funding to help finance the projects of the centers, in which the research of doctoral students would play a major, though not exclusive, role. The principal function of the NCTE would be to publish or otherwise disseminate findings from the research of the centers. A complex of such centers might go far in reducing fragmentation of research in the field, as study could build on study, and would make possible the pilot studies and trial runs necessary for significant research. Centers might address such large-scale problems as the effect of dramatic activities or of a drama-centered curriculum on learning in English and on personality, response to literature, structure of an English curriculum for multi-cultural and multi-ethnic groups of students, response to non-print media and their effects on learning in English.

TWO SUGGESTIONS ON LONGITUDINAL STUDIES

Reflecting more or less incidental references in the preceding section are two suggestions for longitudinal studies, which may be possibilities for research centers or for individual researchers who can find the wherewithal, financial and otherwise, to carry them out.

Developmental patterns of response to literature

A longitudinal study, spanning the nine or ten years between the third or fourth grade and graduation from high school, would be an application to literary response, its problems and manifestations, of the kind of study Loban has done in language development. The purpose of the study would be to chart the growth of response to literature in general and specifically to map the development of responses to such literary elements and techniques as irony, allegory, plotting techniques such as stream of consciousness, and so forth. Such a study would have important implications, obviously, for curriculum planning in literature. A pilot study which may lead into a major longitudinal study of the type just suggested is now underway under the direction of Alan C. Purves and the sponsorship of the Research Foundation of the NCTE.

Dramatic activities

Like the first, this suggestion gives special emphasis to one made earlier in this paper. The value of dramatic activities such as mime and improvised drama in the English curriculum needs test in a longitudinal study of seven or eight years to detect effects on such skills as reading of literature and writing as well as on such general factors as ability in creative thinking and development of personal image.

TWO SUGGESTIONS ON CASE STUDIES

Since the Anglo-American Seminar on the Teaching of English in 1966, many authorities have been telling us that we are in the era of a "process-centered" curriculum in English. Yet the profession at large knows little about the major processes that lead to effective learning or performance in English. Had case-study technique been more in vogue in the research of the last fifteen or twenty years, we might be in a better position to talk sense about a process-centered curriculum than we actually are.

Response to literature

Careful case-study research combined with the kind of longitudinal study suggested above might produce a much clearer picture than we now have of the literary experience

and its development. Case studies of students' engagement with poetry, fiction, and drama, which could chart responses of the student while he is reading a selection, immediately after the reading, and long after the reading, might go far toward clarifying the rewards which different kinds of students derive from different kinds of literary material and the difficulties which the readers experience. Obviously, a number of studies are needed, each dealing with different kinds of students and different kinds of literary material.

The composition process

The possibility of follow-ups to Janet Emig's exemplary study was cited in the preceding section. Case-study research might be directed very profitably to the process the student goes through in "translating" thoughts framed in oral language to clear and effective writing and to the obstacles students may encounter in this process. Robert Zoellner's "talk-write pedagogy" is directed, for example, to this process, but his theories need rigorous testing.

TWO SUGGESTIONS ON IMPROVING EXPERIMENTAL STUDIES OF TEACHING METHOD

Experimental research may have waned in importance and built up a climate of suspicion about the genre, but further experimental studies in which teaching method is the variable need to be carried out. The following are two ways in which such studies might be improved.

Data on classroom life during the experiment

Most experimental studies of teaching methods could be much more illuminating than they usually are, whether or not the final verdict is "significant" or "not significant." Typically, such studies feature pre- and post-measurement of the effect of the teaching method. Often, description of the method which is the variable and of the control technique is perhaps the only information on what went on in the classroom during the days, weeks, or months of the experiment. Systematic gathering of data, day to day, on teacher and student behavior might give real clues to the overall problem of methods: what kinds of students learn what kinds of behaviors or are motivated or "turned on" by what kinds of teachers? The chapter on "Measuring

Classroom Behavior by Systematic Observation" by David Medley and Harold Mitzell in the *Handbook on Research on Teaching*[54] might be a useful guide for planning how to gather the needed data.

More precise statistical analysis

The refrain, "no significant differences," might have been altered in many experimental studies if appropriate statistical analysis had been used. Usually studies of teaching method in English are concerned with behaviors that develop very gradually, and the typical study, running for a school year or less, produces slight gains on the criterion measures. When the effect is small, the number of subjects has to be large. Many studies, in view of their small numbers of subjects, have used statistical analyses which do not have the power to detect small gains. James E. Brewer, Professor of Education at the Florida State University, made available to the writer an unpublished paper which suggests that researchers use the tables in Jacob Cohen's *Statistical Power Analysis for the Behavioral Sciences*[55] to calculate the minimum sample needed for attainment of a fixed level of significance (.05 or .01, for example). Brewer points out that if the sample size is fixed and limited, the power of the statistical test can be calculated beforehand.

CODA With the symphony of duets in Part III concluding this paper, the writer confesses, as a brief coda, that in preparing the paper he was never totally positive of what he was trying to accomplish. Now that the paper is finished, he is not sure what he has accomplished! If research in the teaching of English is a troubled dream, then producing a brief overview of it that makes any specific sense at all is a nightmare of sins of omission and commission. The writer can find safe harbor, perhaps, in the hope that the paper will "provoke discussion." But maybe it will only provoke!

[54] Op. cit., pp. 247–328.

[55] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences,* (New York: Academic Press, 1969).

Anthropological Research Models

CAROL TALBERT
Syracuse University

The emergence of aggressive and nationalistic Third World countries makes it imperative that a non-ethnocentric attitude be transmitted in our schools. The discipline of anthropology views the configuration of norms and values of groups as integrated and adjustive within their own environmental setting.

Anthropology—the study of man—differs from sociology, history, and psychology primarily because of its emphasis upon the field experience and its cross-cultural theoretical approach to social phenomena. Contrary to animals, man has the capacity to create diverse cultural patterns. Domesticated dogs, whether Alaskan or Manhattan bred, all behave essentially the same way—like dogs; whereas Eskimo food preferences and preparation gives us little indication of the food getting habits of the urbanite in Manhattan. Culture is that package of attitudes and customs which people possess and animals do not. Each culture is carried around in the head of each person within it. A man's culture explains to him his relation to the universe, time, other men, nature, and himself.

ANTHROPOLOGICAL PERSPECTIVES ON EDUCATION

Anthropological analyses of educational institutions and personnel within them utilize research strategies and theoretical paradigms stemming from the field work experience and cross-cultural theory.[1] An anthropologist doing field work desires to become totally involved in the group he is studying. In a different culture, it is as if he were a child,

[1] c.f., the series of case studies edited by Spindler and Spindler entitled *Case Studies in Anthropology and Education,* (New York: Holt, Rinehart, and Winston).

though, because he is an adult he must unlearn his familiar ways and learn new ones. A child learns to speak by listening, making mistakes, slowly internalizing grammatical rules and achieving a competence which enables him to predict the actions of others. The researcher slowly and painfully comes to view the natives' culture as they see it themselves with its peculiar contradictions and prescriptions.[2] The ideal way to do anthropology in schools would be to become a student, sit all day, take orders from the teacher, visit other students in their homes, and play with them. The ideal way to study the culture of teachers would be to be a teacher oneself. Yet, as researchers, we cannot be children nor well-socialized teachers. The anthropologist living with Indians on a reservation knows fully well, as do the natives, that he is only there for a while, can never be one of them because he belongs to a different society. Furthermore, he is usually a member of the dominant white middle class higher in social status than those he is studying. The alternative is to be a tolerated onlooker, participating as much as allowable in everyday customary activities. We can ask many questions, follow people around, become a part of the scenery and hopefully be taken for granted. We can hope that people cease structuring their interactions to fit our expectancies.

Analysis of children's attitudes is doubly difficult because of their verbal limitations. Observation of them and participation with them in their homes is a necessity if we are to comprehend their actions in the classroom. For example, in order to understand the attitudes of young children toward their teacher we must know their perceptions of their own mothers.[3] A Puerto Rican boy sees adult women as threatening and has seen other adult males resent and combat the controlling mechanisms they exert. An adult authority, having the same sexual identity as his mother, will have a different meaning to a middle class boy; He has a fairly close supportive relation with his mother which easily carries over into the classroom with his female teacher.

[2] c.f., M. Froelich, *Marginal Natives. Anthropologists at Work*, (New York: Harper and Row, 1970).

[3] c.f., E. Fuchs, *Teacher Talk*, (Garden City, N. Y.: Doubleday & Co., 1969).

Field work gives the meaning, from the children's point of view, of the children's actions. These meanings are not on a conscious level and to expect to glean them from interviews and questionnaires is simplistic.

THE COMPARATIVE NATURE OF ANTHOPOLOGICAL STUDIES

Anthropological analysis of education is a comparative science.[4] Development by the researcher of categories of experience and activities explains commonalities of educational phenomena in traditional and complex stratified societies. Basic to a cross-cultural view of education is the nature of the transition from childhood to adulthood. The appropriate adult status is clear to the young in some societies: ceremonies make it clear to child and community that the individual be given all the rights and responsibilities of an adult, though we only appreciate the importance of these rituals when we compare this situation to the gradual child maturation in the United States where there is no clear definition of what constitutes adulthood.[5]

Values taught on a mother's knee and values extolled in school frequently conflict in modern societies. In traditional societies there is high concordance between childhood training and adult responsibilities. The child wants very much to be an adult and usually adults are happy to pass on their skills to the young.

North American Indians for example, learn in their homes to be cooperative, to suppress their individuality, and to be non-competitive.[6] This socialization pattern was highly adaptive to tribal life as the survival of the group depended upon individuals who could work together. Frustration and failure have come to Indian children with the advent of Western education with its promise of opportunity based upon an individualistic achievement antithetical to Indian values.

Social complexity need not generate failure, as discussion of education of Communist China reveals the concordance

---

[4] The following discussion is indebted to the suggestions of J. Singleton, "The educational uses of anthropology" in T. Weaver, ed. *To See Ourselves* (Glenview, Ill.: Scott, Foresman, 1973).

[5] See, for example, J. Middleton, *From Child to Adult,* (Garden City, N. Y.: Natural History Press, 1970).

[6] H. F. Wolcott, *A Kwakiutl Village and School,* (New York: Holt, Rinehart, & Winston, 1967).

of home and school training.[7] Hard work, egalitarianism, moderation, and cooperation are stressed in both familial settings and ideological tracts. In school the young child is part of a group, his fortunes follow that of his group, and individualism is discouraged. This educational ideology serves to enculturate the child with the values of Communist China.

Cultural comparison leads us to understanding of cultural diversity. We see the wholeness of each culture. It also helps us to see the vicissitudes of the institutional bureaucracy in the United States which seems to encourage children of the upper and middle classes to succeed and the remainder to expect failure. Consider the frustration which must be felt by many teenagers in the U.S. public schools who are treated as immature and stupid and yet when outside infantilizing classroom walls exert authority in the neighborhood and family as full adults. Such is the case with many adolescent minority and white children. These insights lead us to question the prevalent functions of prolonged adolescence and compulsory schooling.

THE RELATIVISTIC VIEW

Stemming from participant-observation and comparison is the third contribution of anthropology, its relativistic view of culture and cultural patterns.[8] We should not view non-Western societies with the implicit standard of our own experiences. We should not use our scales to describe another society. We are not justified in using description which leans on a measurement of the amount of difference between their patterns and ours. This would be akin to describing Chinese language in terms of English grammar. The research of educational psychologist Arthur Jensen for example uses models of intelligence based upon non-anthropological conceptualizations: a) there is a biological base for cognitive abilities which can be measured in intelligence tests, whose validity is based upon ability to predict the functioning of children in schools; and b) the existence of genetic isolates within complex societies which accounts for the lack of 'g' type abilities in Black

[7] Jan Myrdal, *Report from a Chinese Village*, (New York: Random House, 1965).

[8] c.f., E. Liebow, *Tally's Corner*, (Boston: Little, Brown & Co., 1967).

children.[9] The relativist's view would be that intelligence is composed of various genetic combinations and that furthermore, culture defines how one's cognitive ability is expressed. Secondly, to accept that there is little or no interbreeding among blacks and whites (terms "black" and "white" are social, not genetic, categories) or that only whites have a certain genetic component is patently false. Studying other cultures enables us to look again at our own conception of "normalcy." Why is it that males and females compete in the classroom, and are separated during gym practice, and while toileting? Why do males in schools carry milk cartons and girls deliver most of the messages? To us this is natural, the "way things should be," to a person from another culture it may seem amusing, or even stifling to the children.

Anthropological research is holistic. In order, for example, to understand the failure of so many Indian school children one would not be content to blame their failure to read upon the difficulties of English language, their inadequate language training in their homes, the inadequacy of the textbooks, or the lack of proper teacher preparation. Though all these factors are important it is also necessary to understand the Indian's own experience. The discontinuities experienced in school could only be understood when we know the common life cycle of the Indian family; the biological aspects of his early life—health, nutrition, and diet.[10] Analysis of all of these, plus a look at the relation of Indians to whites in a dominant white society would be necessary to explain the failure of Indian schools to ensure the success of their students.

CULTURE AS A DYNAMIC CONCEPT

The concept of culture is not a static term; neither are the everyday patterns of living in quiet equilibrium. Although early scholars looked upon culture as having an existence of its own in a deterministic fashion, we now see culture is defined by looking at how people use it, how it adapts to changing social and physical environs. People within a culture adapt and utilize new resources thus

[9] A. Jensen, "How much can we boost I.Q. and scholastic achievement?" *Harvard Educational Review*, 1969, 39, 1–123.

[10] c.f., E. Erickson, *Childhood and Society*, (New York: W. W. Norton & Co., 1950).

contributing to change and modification. Some elements of a culture will exhibit more resistance than others. There is a core socio-economic adaptation in any society whose essential structure will change very slowly depending upon large scale shifts in the external situation. Just as language changes through mobility, migration, conquest, and technological inventions, its core, its deep structure and grammatical rules will remain fairly constant. English language has adapted to the space age, has borrowed terms from surrounding nations, and also has retained its core grammar. Children and teachers also adapt to each other. We should not view the school as a one way transmission, the teacher will also react to her students. A child is not merely a passive recipient of the culture of the school but one who, by his actions, is also adapting and exerting influence over that institution.

The anthropological perspective views social systems as aggregates of institutions which integrate and adjust groups interlocks through economic, religious, political, and social networks. All the groups in a society do not stand in equal relation to each other and institutions which serve one group's needs may not meet the needs of another. Each individual in a social system fulfills the duties attendant upon his particular status as well as acting in different ways to other individuals under and over him in the social hierarchy. For example, an excellent way to understand an urban school principal's relations with his teachers, parents, and School Board administrators would be to follow him in his daily duties for several months. We would then come to know what he does every day, what are his principle concerns, what are his chief frustrations. When studying institutions of education we are also studying social change agents who in many situations are striving to transmit mainstream values to children whose experiences have taught them that such goals are unattainable. Teachers or principals may be marginal to some extent to both the institution and to the ethnic group of which they are a member.

In summary, anthropological research can enable one, by viewing other social systems, to rediscover the arbitrariness of one's own value system. Those who want to apply anthropological principles to their own particular research may feel

it is unduly difficult. The following are suggestions for utilizing the principles enunciated above within a limited scope.

ANTHROPOLOGICAL RESEARCH MODELS

Culture Through Language

In order to decrease the subjectivity of the researcher, one must discover the labels with which the subjects view their world. Linguistic research has been particularly successful in enabling researchers to elicit the informant's own categories. A definition of culture is that knowledge a person needs to utter grammatically correct sentences—what one must know in order to communicate with members of that culture. We are asking "what do those people see themselves doing?" We must view our subjects as the experts and ourselves as the naïve students. Anthropological linguists have carefully studied, for example, kinship nomenclature among societies and organized the innumerable ways in which *relationship* can be viewed. Aside from the biological fact of maternity, the manner in which responsibility and relationships are reckoned is arbitrary. That is not to say that universal principles such as incest taboo, responsibility for enculturation, license with some relatives, extreme formality and respect with others will not be found. The point is that the particular people considered "relatives," those given respect, these responsible for daily infant care—cannot be predicted with certainty from one culture to the next.

Discovery of Cultural Patterns

The first task of a researcher is to find an appropriate locus for his study.[11] There are many places in and outside of a school where learning occurs. Depending upon our particular interest we might wish to sit for periods of observing in a classroom. We might give the children a tape recorder to obtain their quiet communications during school. Playgrounds, washrooms, and hallways: all would be fruitful cultural scenes. An ethnography of schools should ideally divide research time into formal observations in the classrooms of particular teachers, interviews with the teachers, observation and participation in the playground, as well as periodic visits to the homes of some

[11] The following discussion is indebted to J. Spradley and D. McCurdy, *The Cultural Experience*, (Chicago: Science Research Associates, 1972).

children for additional observation and informal interviews.[12]

After one has selected one's locus it is important to look for one or two key informants. Cultural patterns and individual and idiosyncratic traits must be distinguished and for this reason it is well to talk at length to two or more people at least. There will probably emerge an individual whom one can trust and who seems to understand what one is doing; such people are indispensible to anthropological inquiry.

In discussing field work, I should also mention culture shock, feelings of anxiety and even depression experienced when one really becomes involved with persons different from oneself. Getting inside the world of children presents an especially difficult problem. An adult would have to learn very quickly to discard the role of authority and refrain from condescension.

Whatever group one is studying, it is very important for one to be with them, chat, observe, and participate as much as possible. After having gained some rapport—trust—it might be possible to use a tape recorder, as it is always invaluable to record actual communication. The people one observes may let one take notes in a notebook, though one should either take notes almost continually or not at all. The impression a researcher gives as he listens then notes down only some items indicates to the subjects what the researcher thinks is important. If none of these alternatives are possible then one will have to rely on taking notes away from the 'field site.'

The question will naturally arise about how many people one must observe and interview. My suggestion would be that at least three or four would be a necessity. If we think of cultural patterns as linguistic patterns then interviewing one person who lisped would not let us know whether lisping was a rule governed by culturally accepted norms or if it was idiosyncratic. One can double check the information an informant gives at a later time, ask other's opinions, and trust one's rapport.

The essence of the anthropologist's task is to discover the categories which informants use. What are their cat-

[12] c.f., C. Talbert, "Studying education in the ghetto" in T. Weaver, *To See Ourselves* (Glenview, Ill.: Scott, Foresman, 1973).

egories and what do they mean? How is their knowledge organized and classified? Answering these questions is what discovering culture is all about.

A helpful strategy for beginning research is asking a *grand tour* question; i.e., have an informant tell about the activities and people one is interested in. This will help one find out the meaning of the different parts of the cultural setting you are studying. An individual's conception of space, time, and activities will be described when the person outlines to you what he does and where he does it. Linguistic taxonomies are ever-increasing inclusive domains of reference, and the researcher can discover which of the categories are within the different, more general ones. We want our informants to tell us that this X is an attribute of Y and so on. To discover the rules for ordinary behavior we should ask structural questions such as "What are the different ways to. . .?", or "What are the different kinds of . . .?" These categories exist in people's minds and they will act upon them as they define meaning.

Though field work would be considered necessary for a graduate in anthropology many of the concepts and methods can be used by non-anthropologists. In a sense we all do anthropology when we find informants around us who explain to us what they see.

ETHNOGRAPHIC RESEARCH AMONG SCHOOL CHILDREN

Language Training

I will briefly discuss anthropological description of minority children in school and home to demonstrate research models. Verbal learning in the homes of poor Black children in a southern town, Rosepoint, has been described in a recent text.[13] The speech of young children was carefully recorded and analyzed along with a chronicle of everyday activities which the author observed. After gaining rapport with her subjects, she was allowed to tape on-the-spot interactions between adults and children. Categorization were made of speech events using a model of restricted (Rosepoint) and elaborated (middle class

[13] M. C. Ward, *Them Children*, (New York: Holt, Rinehart, & Winston, 1971).

speech.[14] The author related the preponderance of the restricted code usage as expressing a basic view of the nature of children in Rosepoint.

> Given certain beliefs about the nature of children and their discipline, the means of control, verbal or not, follow logically. . . . children are unable to negotiate or bargain about their position vis-à-vis authority.
>
> An alternative . . . parent-speech . . . (from the) American middle class, is the practice of "casuistry," the application of general ethnic principles to specific cases.[15]

This analysis is of language function—the uses to which speech is put—in Rosepoint. Types of categories of language use are: corrections, requests for information, strategies of control, threats, child manipulation of parents, communication of affection and aggression. These language behaviors are explained as adaptive to the environment of rural Rosepoint with its close informal in-group interaction, as well as the historical experience of slavery. In her conclusion, the author points out that language behavior in the home is not the kind rewarded in the classroom.

The cultures of home and school are described by the author as conflicting. She relates the high dropout rate, low literacy, and resultant abiding poverty of Rosepointers as these are related to the unnatural environment of the schools. The Rosepoint pre-schooler determines his own schedule for eating, sleeping, and playing. The context of his play is unsupervised and depends upon his imagination. In Rosepoint, communication is verbal rather than written and replete with extra-verbal communications. To members of a culture rich in in-group lore and oral traditions the written word is a pallid substitute. Whatever the reason, Ward notes a child, unconvinced of the need to commit his thoughts to paper, is primed for trouble at school.[16] Ward's description, based on her participation and observation, augmented by the tapings gives us a view which is holistic and relative. There are limitations

[14] B. Bernstein, "Elaborated and restricted codes: their social origins and some consequences," (A. Smith, *Communication and Culture*, (New York: Holt, Rinehart, and Winston, 1966).

[15] Ward, *op. cit.*, p. 74.

[16] *Ibid.*, p. 91.

in her conclusions; as she does not differentiate between social class and ethnicity and does not rigorously exemplify and define the middle-class speech which she compares with Rosepoint speech.

Intensive Family Observation

In-depth observation and participation is discussed by Jules Henry in his study of the families of psychotic children.[17] The author lived for a week at a time in the homes of four hospitalized children, chatting, helping with household chores, or joining them in excursions. Feeling it was intrusive to take notes while in their presence, he would periodically retire to his room to record his observations. Henry is primarily concerned with the extent to which common cultural attitudes and behaviors become distorted or exaggerated in families. An example of the violence of American culture is recorded in a scene in which a father playing with his young son becomes increasingly rough and ceases only when the play becomes overtly violent and frightening to the child. Concentrating upon the ordinary, the author invests it with meaning and deep symbolism. Ritual, for example, is described as essential to a child's growth. Repeated patterns of care and nurture which can be depended upon are comforting to a small child. By ritual is meant such ordinary customs as wrapping a child in a warm cocoon of a towel after a bath, rubbing his body, and crooning to him. Established rituals, when neglected by parents, give little structure to a child's existence.[18]

Cultural Traditions as Data

Anthropologists have always relied upon folk tales and oral narrative to add dimension to their accounts of tribal origins and beliefs. Black-American society has a rich heritage of oral stories and tales. The toasts, jokes, and repartee of urban Black men give us insight into the conflicts and pleasures of a social group. Narratives, plus knowledge of social and economic circumstances, have

[17] J. Henry, "The naturalistic observation of the families of psychotic children," *Proceedings of the Third Institute on Preventive Psychiatry of the State University of Iowa*, (Iowa City: State University of Iowa, 1961).

[18] c.f., J. Henry, *Pathways to Madness*, (New York: Random House, 1971).

been interwoven into a total cultural descirption by folklorists and anthropologists.[19] The oral themes can be the data for a psychological perspective of the dilemmas of poor Black men. Thematic structures stemming from African folk tales and from the time of slavery can be found in current usage. The conflict expressed by the men about their authoritarian mothers and exploitive wives gives insight into the attitudes young Black boys must hold towards their mothers and female teachers.

Language or Dialect, Culture or Sub-culture

William Labov taped the casual speech of a number of young Puerto Rican and Black gang members by hanging lavalier microphones from each of them.[20] There were various language styles; narrations, "playing the dozens," jokes, tales, and angry outbursts. Rather than describe these utterances by their level of approximation to Standard English grammar Labov analyzes them in terms of their own structural and phonological rules to construct a phonological rule-governed system of Black English. Labov also presents a structural analysis of the rules for oral style. These narrative rules are distinctly different from the written language found in elementary school primers. The language spoken by Black-Americans is not a substandard dialect of standard English, just as their culture is not a "sub" culture.

Cultural and Linguistic Diversity

Linguistic analysis also lends itself to study of relationships between groups in a complex society.[21] Language variation, by speakers in everyday situations, can indicate their interpretation of the meaning of the social situation—of the status of the speakers in relation to the listener. The lack of mastery of formal Standard English can stigmatize some individuals. Mainstream language forms are seen by most members of a society as the right and proper way to express oneself. Those who do not use Standard

[19] R. Abrahams, *Deep Down in the Jungle,* (Chicago: Aldine Publ. Co., n.d.).

[20] W. Labov, *A Study of the Non-standard English of Negro and Puerto-Rican Speakers in New York City,* Cooperative Research Project #3288, Vols. I & II. (New York City: Columbia University, 1968).

[21] J. Gumperz and D. Hymes, eds., *Directions in Sociolinguistics,* (New York: Holt, Rinehart, & Winston, 1972).

English are thought less articulate and intelligent. Adherance to the moral and aesthetic value of Standard English is linguistic imperialism.

All of us switch unconsciously between standard (formal) and non-standard (informal) forms frequently. Speakers of Black English also shift from informal to formal, but the switching from Black English to Standard English is more than a style shift; it is a grammatical shift. Studies of bi-lingual speakers indicate that personal and emotive utterances are in native language whereas Standard English is used for communication with non-natives and in formal situations. Use of Spanish, for example, can communicate togetherness among Puerto Ricans. Harsh and aggressive ghetto expressions by militants can cause irrational fear when heard by middle-class listeners. These mannerisms and grammars are impossible for a white speaker to imitate and further demonstrate his separateness from Blacks and promote their solidarity.

Teachers in public schools appear to misunderstand the nature of language and language diversity. They teach English grammar as if it were totally rational as a system rather than a conglomeration of borrowings and inconsistencies. They express dismay at the utterances of a Black child whose Black English grammar is thought to reflect his intellectual and moral "laziness." The rapid-fire speech he hears in his home, replete with subtle intonations and rhythms is discouraged and the child is taught in the expressionless flat monotone of the middle-class speaker.

The Native's Point of View

Linguistic research can assist us in discovering the meaning of the educational process of the student. Linguistic analysis of the child's perceptions can provide insight into what is really being learned in school.[22] Spradley's text illustrates ethnographical research in a junior high school by interviews with students in a casual setting.[23] We are given a description of "what teachers do . . ." in a taxonomy, with its various attributes, listed within it. Below is an

[22] c.f., C. R. Cazden, V. P. John and D. Hymes, *Functions of Language in the Classroom*, (New York: Teachers College Press, 1972).

[23] Spradley, *op. cit.* pp. 110–111.

illustration of "what teachers do" in school and the manner in which it is graphically presented.

Figure 1
Taxonomy of the Domain of "People at Midwest Junior High"

Faculty		
	Teachers	Shop teacher
		Gym teacher
		Core teacher
		French teacher
		Special teacher
		Science teacher
		Math and science teacher
		Art teacher
		Music teacher
		Home Ec. teacher
		Homeroom teacher
	Head Faculty	Principal
		Assistant principal
		Counselors
Subs		
School Nurse		
Cleanup Men	Janitors	
	Dirty towel man	
Kids	Troublemakers	
	Goody-goodies	
	Brains	
	In-betweens	
	Cool kids	
	Sads	
	Colored kids	
	Deaf kids	
	Loners	

PEOPLE AT MIDWEST JUNIOR

As the author points out, the information in this paper is only a portion of the folk knowledge that the eighth grade girls we talked with have in their minds.

What is presented comes from their own point of view. It is a beginning of a study of their culture, which is important

> to a study of urban education . . . at Midwest Junior High School meant to the children. There is a pattern of teachers picking on kids and kids picking back and acting up. There is very little cooperation between them. The diversified population of this city is also reflected in the shallow judgements which the children make of one another based upon their dress, looks, and manners.[24]

We can see from looking at the graph that the important factors for example, in the students perceptions of the "colored kids" are their personal feelings, "coolness," intelligence, and how much the kids talked.[25]

Cross-cultural Studies of Education

Cross-cultural analyses of educational processes are concerned with such basic questions as the nature of teaching, modes of control, and status of the teacher. Anthropologists have devised various organizing systems by which the raw data from observations in a classroom, or any educational setting, can be categorized and systematized. I shall discuss one example, of twelve major sections each divided into many subsections, with an explanation of the outline and a theoretical orientation to the problems of education in its cross-cultural aspects. Below is a sample from one section concerning the coding of educational process.

I. *On what does the educational process focus?*
   a. values. b. value conflict.
II. *How is the information communicated? (teaching methods)*
III. *Who educates?*
IV. *How does the person being educated participate? (What is his attitude?)*
V. *How does the educator participate? (What is his attitude?)*
VI. *Are some things taught to some and not to others?*
VII. *What limits the quantity and quality of information a child receives from his teacher?*
VIII. *What limits the quantity and quality of information a child gives a teacher?*
IX. *What forms of conduct control (discipline) are used?*
X. *What is the relation between the intent and the result of education?*

[24] *Ibid.*, p. 117.
[25] *Ibid.*, p. 109.

XI. *What self-conceptions seem reinforced?*
XII. *How long does the process of formal education last?*[20]

Subsumed under these categories are many illustrations and explanations of these aspects, but basically the same listing and cataloguing of attributes is made, although more complex and definitive than Spradley's model.

A recent study of New York city schools used many modes of gathering information: observers in the classrooms, interviews, teacher ratings, and examination of textbooks. On the basis of their observations of teachers and students talking with each other the author developed various codes of the many behavioral types. An illustration of her interaction coding follows.

1. *Neutral acceptance or affirmative*: Either no answer, or a comment like "yes," "that's it," "right," delivered in a neutral tone.

2. *Positive evaluation*: "Good boy," "very good," "very nice," "that's fine," or more extensive remarks, such as "Good, you're working on it," "Good, now you've proved that." Remarks such as "good," though coded as positive, present a problem in that they may be delivered automatically as virtually neutral acceptance or said with an added emphasis and supportive intent. Positive comments need not only follow correct answers; remarks like "Good, that's part of it," or "Good, you're trying," can follow mistakes.

3. *Negative evaluation*: "You don't know too well. . .," "Stick to your topic," "We're not talking about that." Negative evaluations need not only follow mistakes. Teachers can respond to a correct answer with, "That's enough for you" (cutting off discussion), "Make it in the form of a sentence," or "Let's label our answer" (if a cent sign, for example, is left out of a mathematics example).

4. *Correction or completion* of an answer, or *passing on* to another child without comment.

5. *Opening responses:* Either following up a correct answer, seeking the errors in an incorrect answer, or affirming and elaborating upon or developing an answer

[20] J. Henry, "A cross-cultural index of education," *Current Anthropology*. 1 (July, 1960).

leading to a new question. Most of the responses that had to be coded as "opening" were on a trivial or superficial level. The better "openers" observed were remarks like "Good, how did you arrive at that number?" "Good, show us on the board," or "All right, why do you say that?"

6. *Miscellaneous:* Other or unclear incidents.[27]

This study of content and structure of classroom teaching is augmented by interviews with both teachers and students. The data question many of the common assumptions of educational researchers. By rigorous observation the writer demonstrates, for example, that permissiveness and high expectation of pupil achievement do not always occur together.

In analysing techniques of classroom management, Leacock states that she has developed three general categories: supportiveness, enabling of independence, and leeway for child interaction. The attributes of these general categories were found in a number of ways. All members of the observer teams independently rated each teacher on a five-point scale. Teacher remarks were coded as positive, negative, or neutral and a count was made of their distribution over a three-hour period. Teacher actions and statements were coded: directives to the class, noise and movement allowed, handling of transitions and routines, structuring of learning situations, assignment of responsibilities, child participation, allowance for free periods. Physical features of the room were also discussed.[28] The three variables were found to be highly interdependent yet there was no direct relation between teacher's supportiveness in the area of behavior and the value or emphasis she places on learning as indicated by the ratings of teacher's interest. Furthermore, there is no consistent relation of "supportiveness" to proportion of negative to positive teacher statements. Leacock then states that her analysis offers alternative "models" to the theoretically ideal climate for learning in one case: the enabling of interchange and structuring for co-operative work relations combined with high standards for learning.[29] The impli-

[27] E. Leacock, *Teaching and Learning in New York City Schools* (New York: Basic Books) p. 66.

[28] *Ibid.*, p. 90.

[29] *Ibid.*, p. 113.

cation to me is that modes of socialization in Black homes may be reflected in teaching style of Black teachers. The author also notes the emphasis upon moral training in social studies. The ideals of our democratic nation are somehow distilled into "being a good citizen" and obeying rules.

Long-term Studies

Let us examine a long-term study of Black kindergarteners and first graders in urban ghetto classes. The study combined observation with ethnography. The first year was devoted to study of classroom interaction. Compilation of negative and positive responses by the teacher based upon the following code was made for three separate months in two kindergartens.

Figure 2

Code of Interaction.

Pupil, whose identity must be recorded, initiates an action; teacher responds. The response is then coded as (+) or (−) in accordance with the following guidelines.

POSITIVE (+)

1. Pupil is engaged in disruptive or disapproved behavior, teacher ignores it, though aware of it.
2. Teacher has asked pupil a question, pupil answers, and teacher in response indicates approval.
3. Teacher has asked pupil a question, pupil answers, and teacher indicates approval by asking for further information on the subject.
4. Teacher expresses approval for an action by the pupil which occurred in the past.

NEGATIVE (−)

1. Pupil is engaged in disruptive or disapproved behavior, teacher indicates disapproval, verbally or nonverbally.
2. Teacher has asked pupil a question, pupil answers, and teacher in no way indicates approval.
3. Teacher has asked pupil a question, pupil answers, teacher points out what is missing or incorrect.
4. Teacher expresses disapproval for an action by the pupil which occurred in the past.

EXAMPLE OF CODING OF INTERACTION.

If a child was disturbed by another child and went to tell the teacher it should be coded in the following manner, depending upon the responses of the teacher.

(+) for Child A, who was disturbed, if teacher praises him for being a tattle.
(–) for Child A, if teacher reacts negatively to him saying, for example, "Well, if you hadn't been out of your seat, David would not have hit you."
(+) for Child B, who was snitched on, if the teacher ignores his behavior or signifies approval.
(–) for Child B, if the teacher reacts in any negative manner.[30]

It was concluded that the total amount of one-to-one interaction lessened considerably as the school year wore on. During the observation periods there were many children with whom the teacher never had any one-to-one conversation, and a selected few children were the focus of most teacher interaction. The teacher appeared to be working toward a classroom in which the total amount of interpersonal communication with her kindergarteners was minimal, with most of the communication being impersonal directives. Furthermore, there emerged a central core group of children in close around the teacher, who interacted with them the most. There was also a peripheral group of children close around the teacher, with whom she interacted less and less. Over time the interaction between members of this peripheral group intensified which made them further removed from actual teaching. In the first grade the amorphous groups were rigidified into hierarchical reading levels, the central children being in the high group, the peripheral in the lowest group.

It appeared from our research in the various schools that early selection by the teacher of the "most likely to succeed" children had an almost unalterable effect upon subsequent achievement. It was not observed in any of the classrooms that a child viewed by the teacher as potentially not successful ever became a "star," the role of "pet" or "star" appeared to follow a few children year after year.

The following year the same teacher was asked to rank the new first graders from "most likely" to "least likely" to succeed. The children's ranking by the teacher closely followed their placement in high, middle, and low reading groups—in the first grade.

[30] C. Talbert, "Interaction and adaptation in two negro kindergartens," *Human Organization* (Summer, 1970) p. 108.

The most favored children were those whose behavior and speech most closely approximated that of the mainstream ideal—cleanliness, verbosity, neatness, "prettiness," or "cuteness" were good attributes. Children who were described as having mumbled speech, a urine smell, sloppiness, old clothes, etc. were the least favored by the teacher.[31]

In order to distinguish the effects of social class and ethnicity, or rather, to understand their inter-relation, I chose a linguistic variable to measure the amount of intrusion of white norms into Black behavior. I hypothesized that those children who were able to switch from Black English to Standard English in the appropriate setting the best, thus exhibiting their middle-classness, would be the ones teacher would invest her energies in the most. I also hypothesized that in certain settings both teacher and pupils would use primarily Black English and in others there would be code switching to Standard English.

The language of the pupils was recorded in three different linguistic tasks; reading, looking at pictures, and chatting (telling scary stories) with each other. I found that all of the children used phonology associated with Black English in the casual settings and that they all used some Standard English forms when engaged in formal tasks (reading). What was most interesting was that the amount of code switching (i.e. child facility in using Standard English) was closely correlated with higher teacher ranking. Furthermore, the teacher who was Black, used Black English forms when speaking one-to-one with her children and Standard English forms when engaged in giving directions and speaking to the class as a whole.

Below is a systemization of the grammatical forms which will be used under certain circumstances if we have knowledge of the Sender, Receiver, Channel, Setting, and Message form. We see from this model that Black English usage is found in particularly well-defined situations which are distinct from those in which Standard English is used. The demands of the classroom are such that Standard English will be used in certain tasks which, are more im-

[31] H. Gouldner, J. W. Bennett, M. Durbin, R. Rist, and C. Talbert, *The Black Child in School* (Final Report. OE #2-6771. Social Science Institute, Washington University, St. Louis, Mo., June, 1971).

personal, pedagogical, and unemotional. Black English is used in situations which are informal, intimate, and egalitarian.[32]

Figure 3

Ethnography of B.E. and S.E. Speech Events: Determination of Code Usage

Code	Usage	Sender		Receiver		Channel		Setting		Message Form
BAE	←	AD	+	CH	+	V	+	1-1	+	IC
BAE	←	CH	+	CH	+	V	+	PG	+	GR,SG,JK,IC
BAE	←	CH	+	AD	+	V	+	1-1	+	IC
SAE	←	AD	+	GRP	+	V	+	CL	+	DIR
SAE	←	CH	+	AD	+	V	+	CL	+	RDG
SAE	←	GRP	+	GRP	+	V	+	PG	+	RDG

Key:

BAE = Black American English
SAE = Standard American English
AD = Adult
CH = Child
V = Verbal
PG = Peer Group
1-1 = one-to-one conversation
CL = Class
DIR = Giving Directions
SG = Singing
JK = Telling jokes, spooky stories
IC = Informal conversation
RDG = Reading

On the basis of these data *Black teacher* can be described as both *Black mother* and *White educator*. The conflict between these two roles is that learning modes prevalent among Black adults toward children is very different than those behind most educational theory. As Black mother the teacher is authoritarian, non-redundant, punitive, and little given to lengthy explanation. As teacher in a school she is expected to be nurturant, permissive, and replete with multiple examples and explanations for the children. Secondly the attributes of Black family life which enable the child to learn how to cope with his environment are not allowed in the classroom, i.e. high interaction with peers, freedom and mobility (physical and verbal), the opportunity to listen to and observe adults in discussion of their problems and solutions. Classrooms commonly are very quiet with children rivetted in their seats and the teacher speaking "at" the children rather than "with"

[32] C. Talbert, "The weeding out process." Unpubl. Syracuse University, Syracuse, N. Y. 1972.

them. Furthermore, as pointed out, the functions of language among poor Blacks is different in some ways from its functions among middle-class persons. Language, or the non-use of language, can have high survival value when a Black is in contact with members of the dominant society. The child often witnesses adults speaking with welfare workers, policemen, or truant officers, and learns to use speech as a manipulation and disguise.

My participant-observation among poor Black families corroborated the conceptualization of culture-specific linguistic function and structure. The children I observed were verbally creative and aggressive in their everyday speech. The forms of verbal dexterity and expertise rewarded are cleverness, intonational control, even the shock value of one's choice of words. In the classroom, monotonic repetition is rewarded. Interestingly, the children who become the "stars" in the classroom were allowed to become verbally freer, more aggressive, and more creative with the teacher. In other words her reward for their conformity was a bit more freedom.

DOING ANTHROPOLOGY

Doing anthropology is more than sitting in a few anthropology classes. Anthropology requires a personal involvement with groups of strange people with unfamiliar customs. I have outlined some methods of obtaining information which ranges from intensive participation to non-threatening interviews, but I want to make it clear that true cultural description demands more interviews. It takes at least a year's field work, plus years of scholarly study, for one to begin to understand cultural patterns. Yet I also want to encourage others to experiment with methods of observation and participation for it is also of utmost importance that those concerned with education be made more aware of linguistic and cultural diversity. It is imperative that teachers appreciate that their ideas and values are not shared by everyone and that just as we hold to our favorite prejudices and stereotypes with irrational fervor so do all other peoples hold attitudes which may even appear contradictory or destructive but withstand all efforts to change them.

Measurement and Evaluation

JEREMY D. FINN
State University of New York at Buffalo

To many, the terms "research design," "measurement," and "evaluation" connote a set of technical skills involved in the collection, analysis and interpretation of quantitative data. The data are usually scores, obtained by measuring the characteristics of pupils in some educational setting (e.g. aptitudes, achievements, attitudes), the characteristics of other individuals within the setting (e.g. parents' educational levels, teachers' expectations), or characteristics of the setting itself (e.g. teacher/pupil ratio; number of texts per pupil; adequacy of time, space and lighting facilities).

There have been a number of advances in evaluation technology which have potential utility in education, and particularly in the language areas. I will describe some of these later in this paper. But there is a related set of more basic issues that needs to be raised if not settled, before we embark upon the technical considerations. For it seems that it is possible to become so enraptured with quantitative tools that many lose sight of their appropriate purposes and roles in the total evaluation process. Evaluation involves a much broader set of activities than the technical or mathematical applications. A close look at some of the non-technical and perhaps more difficult evaluatory aspects of our research, is necessary before we can decide what technical tools are, or are not, appropriate to our needs.

EVALUATION

Evaluation in its broadest sense is the judgment of value or worth of an object, event, or idea, according to one or more criteria. Typically in conducting research or evaluation studies we "evaluate" in three stages. *First,* we

identify some problematic situation which we deem worthy of study. The problem may exist in a theoretical context, as, for example, the lack of adequate knowledge concerning the effects of educational environments upon the individual. Or it may exist at a more immediate level. For example, we may wish to understand the effects of introducing reading through ITA, a curriculum currently in popular use.

Second, we choose an approach to the study of the problem. In empirical work, we decide upon our starting points or assumptions about how people behave. We may select or construct a theory, a set of hypotheses, or a set of propositions about the forces effecting the outcomes of concern. From these we decide the sorts of variables, measurements, and observational units (e.g. pupils, classes, teachers, homes) upon which to focus our attention.

Finally, we evaluate when we analyze and interpret our results. (Note by comparison that measurement—the assignment of numbers to individuals according to their responses to a given set of stimuli—is not in itself evaluation). We select a set of analytic tools, hopefully designed to answer the major questions of concern. We review the quantitative outcomes and forward interpretations in terms of the variables we have been studying. We draw conclusions and propose recommendations which, according to our judgment, seem appropriate. The recommendations may be for additional research, as well as for policy decisions to be based upon our findings.

Each of these three evaluations should be carefully considered. Each can play a significant role in shaping current and future research.

RESEARCH PROBLEMS IN EDUCATION

As recently as fifteen years ago, education could have been regarded as only a means for perpetuating the established socio-political system and its traditions. A fixed set of intellectual attitudes and skills were passed from generation to generation. Youngsters were informed of the intellectual and social limits to which they must adhere. Research in education could assume no charge other than to identify the most efficient means through which these goals may be attained. Focus centered upon problems of class size, textbook organization, and the efficiency of

related pedogogical devices. Achievement was "measured," as the attainment of satisfactory levels of performance, on norm-referenced cognitive achievement tests. The number of content areas was limited.

While this characterization is extreme, and while all may not accept it wholly. I think we will agree that the situation is dramatically different today. The federal government and citizenry expect the educational sector to produce scientists to compete with those of other developed nations. Education is seen as a means toward overcoming problems of poverty,[1] unemployment, illiteracy. Parents not only expect schools to provide basic cognitive skills, but also to prepare youngsters to deal with factors in their personal lives (e.g. sex, drugs, social and political idealism, disease). Students themselves are dissatisfied with the cognitive outcomes of more traditional educational methods, as are teachers. They seek instruction in inquiry processes. They wish to investigate the widest possible ranges of alternatives to the current social and political status-quo, and to the social-injustices they see being perpetrated. *All* question the negative attitudes toward learning, toward society, and toward education, which seem in recent years to have been a more potent outcome of twelve years of schooling than the cognitive skills themselves. And to everyone these dissatisfactions become accentuated by the rapid publication of news events, of educational successes and failures, and of social criticism.

I will not attempt to assess these expectations, nor do I feel that the hopes for improvement are all realistic. But focus upon education in the United States carries with it a message for those of us conducting research and evaluation. Our work is of the greatest potential import and cannot be conducted in isolation. Through concern as well as through public press, we are rarely out of the view of the remainder of American society. We cannot continue to perform our studies of class size, of textbook organization, of correlations of "un-relevant" tests, or of teaching method one compared with teaching method two, with no basic characterization of either method. We

[1] At least until the appearance of C. Jencks *et al. Inequality: A reassessment of the effect of family and schooling in America* (New York: Basic Books, 1972).

must make every effort to evaluate the scientific *and* practical significance of our research efforts. Our research should be not only "interesting," but necessary! And we must strive toward a broadened research perspective, to parallel the expanded roles of the educational system.

The question of research significance is both extremely difficult, and one which is frequently avoided. "The importance we place on our research findings and their explanations and implications determines what we do with our discoveries."[2] Scientists will often disclaim competence or interest in making such value judgments. And none of us really wants to tell his colleague that there's something more important in his work than he seems to realize. Yet the recent arguments of Scriven and Harrison (among others) that we cannot afford *not* to consider the issue merit review.[3]

THREE CRITERIA OF SIGNIFICANCE

Obviously no single definition of problem "significance" will suffice. Instead consider the following three general criteria, posed as questions, which may form a basis to construct a further rationale of quality:

Basic Variables

Are the characteristics of the program we are studying basic to the problem, or are they symptomatic, and perhaps transient and unreplicable?

Example: The effectiveness of any "curriculum package" is a function of the quantity and quality of pupil materials, the degree of teacher preparation necessary for their dissemination, as well as the quality of teaching guides. Simply to compare the effectiveness of the package with some other unspecified curriculum, leaves no clue as to whether the *separate qualities of the package* were or were not effective.

If a comparative study fails to demonstrate effectiveness, it is nevertheless likely that some of the program's

[2] F. I. Harrison. "The process of conducting research: procedures and proceedings." Paper presented at the American Personnel and Guidance Presession on Research Supervision and Consultation, Chicago, November, 1970, p. 4.

[3] Harrison, *op. cit.*; M. Scriven. "Student values as educational objectives." *Proceedings of the 1965 Invitational Conference on Testing Problems* (Princeton: Educational Testing Service, 1966). M. Scriven. "The values of the academy," *Review of Educational Research*, 1970, 40, 541–550.

qualities were effective while others were not. The overall gain, as a combination of many factors, may appear nil. Information about basic qualities provides data for the "formative evaluation" of the materials, culminating in a revised curriculum. Through successive improvements and re-assessments, the effectiveness of the materials may be enhanced until an acceptable level is reached. If the program or package *is* effective, the qualities that make it so may be abstracted, and incorporated into other curricular materials, in the same or different content areas.

The comparison of teaching methods or curricula which embody many characteristics are contingent upon the particular programs being compared. For example the "traditional approach" or the "control treatment" may itself vary dramatically from one administration to another. We must not have the effectiveness of our experimental program be contingent upon such transience. It is for this reason that Lever Brothers (manufacturers of Pepsodent) will not consider Crest (Procter and Gamble) an adequate comparison. Instead they will compare treatments in terms of the quantity of specific additives.

Progressing from general program descriptions to the definition of basic variables is not a simple task. Characterizing curricula, teaching methods, or educational programs generally, in terms of a series of basic parameters, is a task for which we still lack the tools. The discovery and organization of curriculum-descriptive terms and parameters may represent one of the most-significant logical and psychological tasks confronting us today.[4]

Measurement Innovativeness

Are we measuring new aspects of the phenomena we study, and seeking better ways to measure familiar variables?

Examples: Perhaps of all research in education, none lacks new measurement approaches as much as research in reading. There now exist uncountable tests for elementary-level pupils of reading ability (including measures such as provided by the WISC), comprehension,

[4] L. S. Shulman. "Reconstruction of educational research," *Review of Educational Research*, 1970, 40, 371–396.

recall, recognition, and readiness. And there also exist uncountable reports of their intercorrelations—under natural conditions, under experimental conditions—for the highs, for the lows—for Chicago, for Minneapolis—for linear relationships, for curvilinear—using product moment correlations, using anova—etc., etc., etc. Perhaps we have so many separate reports because of fear of extrapolation from one test or one situation to another. Or perhaps we simply wish to replicate, to be certain. This may be one area in which we should disregard the universal recommendation; perhaps "further research is *not* needed."

It is not my intention to disparage all reading research, as there are a number of exceptional efforts currently underway. However there are also a number of ways to approach the measurement issue, which may assist in alleviating redundancy in language-related research. Two of these approaches follow:

First, we may examine the *context* which gives rise to our data. For example, elementary-level children will respond differently to communications with different contents, as a function of their own interests and attention levels. There undoubtedly exist data for the test of this hypothesis (and if not, they can be simply collected). This assumption tells us that both "content area" of the reading material, as well as the "instructional approach with respect to attention demands," must be assessed in attempts to understand reading achievement. With these contextual measures as co-variables, we would expect to gain in our prediction of reading outcomes beyond that provided by aptitudes and readiness.

Here the context is defined by "instructional process variables"—descriptors of the processes embodied to differing extents in any pedogogical approach. Similar process variables have been used to characterize the aspects of home environments which promote cognitive achievement. Davé observed correlations up to .80 between achievement and measures of home environmental processes (e.g. achievement press, language models, intellectual activity in the home).[5] The process-variable ap-

[5] Davé, R. H. "The identification and measurement of environmental process variables that are related to educational achievement." Unpublished doctoral dissertation, The University of Chicago, 1963.

proach to describing educational contexts would appear to have high analytic power.

For a more detailed description of the learning context, we may also wish to obtain both "high-inference" and "low-inference" measures of teacher behavior.[8] Again we can see the need for a logical analysis of instruction to help identify and name the contextual parameters of greatest consequence.

Second, we may analyze the learning criteria assumed to be important. In particular we might examine the *range of anticipated outcomes.* For example, communication through writing continues to be of import in western cultures. Yet our definition and measurement of writing skills, either directed or creative, remains in a most primitive state. To some extent teachers are turning away from the narrowly-defined objectives of correctness in punctuation, spelling, and grammar. But the more complex skills still lack adequate measurement, perhaps as a function of lacking adequate definition.

Consider instead the different sorts of objectives and definitions which may be relevant. For example, is it possible that the outcome variable we should consider is simply the ability to transmit a message, and to do so whenever necessary? Certainly there are others. The technical measurement question becomes one of how to obtain valid indicators of these alternate outcomes. I will leave this question for you to answer; it is far from an insurmountable problem.

The range of outcomes may also be viewed on a time dimension. For example *criterion-referencing* is a measurement trend which is currently attracting much attention. The term refers to an evaluation procedure whereby criteria for satisfactory performance on a specific set of skills, are fixed in advance of a course or unit. Each pupil is rated on his attainment of those criteria, rather than by comparison to other pupils (i.e. norm-referencing).

[8] B. Rosenshine. "Teaching behavior related to pupil achievement: A review of research," I. Westbury & A. A. Bellack (Eds.), *Research into classroom processes: Recent developments and next steps.* (New York: Teachers College Press, 1971).

Ideas of criterion-referencing are as old as the concept of educational evaluation itself.[7] By comparison their implementation has been slow in coming, and is attempted for only the simplest of skills.[8] It would seem useful to expand upon these applications, especially in language-related areas. Language skills are basic to all we do, say, and learn as children *and* as adults. Ways may be sought to measure the attainment of long-term language goals, not just those of specific courses and units. To accomplish this measurement, we need to specify the necessary adult communications skills, and to distinguish them from those which are not essential. We must also develop the technical means of measuring progress at each stage toward those long-range objectives. These technical advances can facilitate school-, district-, and state-wide program evaluation—those concerned with assessing the effects of an eight-year or thirteen-year school curriculum.

Knowledge gained

To what extent will new knowledge be gained through our research?

In some sense, all research produces "new" knowledge, while in another sense no research effort is completely unique. Yet we may ask whether there are means for raising the potential contribution of any study to knowledge in a problem area. That is, how may we increase its "academic payoff"?

We all recognize a need for "replication studies." Yet we must also question the point at which replications or "almost-replications" no longer alter our conceptions about learning. In particular we need ways for assessing the current stage of knowledge in a problem area, and determining what sorts of studies will move us ahead.

With the plethora of names for the same or similar variables, this task becomes even more formidable. Frequently another investigator has researched *essentially the same* problem as that we are concerned with, only in some

[7] E. R. Smith and R. W. Tyler. *Appraising and recording student progress* (New York: Harper & Bros., 1942).

[8] R. Glaser. "Adapting the elementary school curriculum to individual performance," *Proceedings of the 1967 Invitational Conference on Testing Problems* (Princeton: Educational Testing Service, 1967), pp. 3–36.

other situation or under another set of variable names. The barrier may be overcome at times by realizing the basic constructs underlying the manifest behaviors of concern, i.e. by explicating our assumptions about behavior and the variables we study. And it may be overcome at times by systematically reviewing, organizing, and summarizing theory and findings in the area, as seen from a number of disciplines. We must *evaluate,* not just review, and ask ourselves what we can add that is unique. Review and summary effort (e.g. *Review of Educational Research,* ERIC) are stepping stones in this direction and deserve our full support.

Chamberlain and Platt suggest procedures for constructing a "logical tree" of knowledge, and for stating critical untested hypotheses.[9] While their papers are based upon research in the biological and physical sciences, they warrant careful review for application to the social sciences. In contrast, Senator Walter Mondale suggests that it is the critical summary of what is known about a problem, and what is not, that is desparately needed for policy-making.[10] Perhaps in some instances this summary alone is the most worthwhile contribution.

I have suggested three criteria that may be considered in evaluating research efforts: 1) The extent to which the research involves basic and non-transient variables, 2) The extent to which the research considers a range of process and outcome measures, and 3) the extent to which the research contributes to the current stage of knowledge. The criteria are difficult to assess, but must be regarded continually in planning research studies. The necessary measurement approaches as well as the means for summarizing what is known about particular problems, are themselves methodologies requiring further development.

[9] T. C. Chamberlin. "The method of multiple working hypotheses," *Scientific Monthly,* 1944, 59, 357–362. J. R. Platt. "Strong inference," *Science,* 1964, 146, 347–352.

[10] W. F. Mondale. "Education and social change: the congressional perspective," Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971.

A MULTIVARIATE APPROACH

All have undoubtedly heard the "three r's" of research—the three universal recommendations. Whether the topic is children's reading, the reduction of illiteracy, developmental language processes, or ways to teach a second language, we seem always to converge upon the same points of agreement. First, the problem is not solved (although it will be shortly) and "more research is needed." Second, a major obstacle to the solution of the problem is that we have not approached it from a sufficiently broad perspective. Thus we need to provide for "interdisciplinary thinking and research on the issue." And finally, since our simpler and more familiar analytic models have not provided elegant problem solutions, "we must employ the techniques of multivariate analysis."

In all likelihood, we have become conditioned to ignore these recommendations when stated by others, and to include them in our own reports as a necessary condition for completeness. In the preceding section of this paper it is suggested that we give servious consideration to the first and second recommendations—not to accept them as given, but to evaluate them carefully. In the remainder of the paper I would like to discuss the final recommendation, the need for "multivariate analysis." This recommendation like the others, should not be given or accepted without careful evaluation.

Multivariate analysis may be conceptualized in two manners. First, it is a way of analyzing behavioral phenomena. It is based upon the realization that almost no human behavior worthy of study has only a single facet; that behind any measurable trait are components which covary, if only partially; that a "better" scientific description of any behavior is derived through some degree of finer analysis. Further, no observable behavior results from a single antecedent. The "principle of multiple causes" is one we confront in all except the smallest analytic units (e.g. the "one-gene, one-enzyme theory" of genetic action).

The second conceptualization of multivariate analysis is fitting a set of algebraic models to situations with multiple outcome variables. I will defer discussion of the algebraic conception to a later part of the paper. Let us first examine

the phenomena we measure to see where the algebraic models may be applicable.

All educational research assumes a set of learning outcomes. Even a laboratory study of minor facets of the reading process carries the assumption that certain reading skills are important outcomes; a study of teacher personality and pupil self-concept carries the assumption that one or both of these is at least partially a result of the classroom interaction; and so on. A second set of variables describes the means by which the objectives are conveyed to the pupils. These may include teacher and pedagogical characteristics which aid, or hinder the transmission of ideas (e.g. length of lesson; degree of structure and organization). Finally there are the characteristics of the pupils which determine the extent of the learnings to be attained. These may include pupils' abilities and interests, age, sex and social status, as well as other personality factors.

Most research focuses upon a small set of characteristic from one or two of these domains. The purpose is to demonstrate that under natural or experimentally-induced conditions, the correlations among them or "strength-of-effect" is of magnitude X. Important variables which are omitted substantially limit the degree of prediction or explanation. Yet integration of a larger number of traits into the study appears either too costly or too complex to permit a resonable analysis.

Educational studies *can* consider the multiple sources of influence and the multiple outcomes of educational process. Our common goal is to understand the behavior of the graduate of a unit, a course, or thirteen-year career in an educational setting better than we do now. To accomplish this, we must explicitly state and utilize the assumption that the educational enterprise has *both* a manifest and latent curriculum[11] and yields *both* intended outcomes, and others which co-occur but may not be anticipated.

The multiple-input, multiple-outcome assumption can be formulated as a model depicting the classroom enterprise. The model is a way of stating what we know or assume about the problem situation. Sound behavioral models may be

[10] B. S. Bloom. "Individual differences in school achievement: a nishing point?" Paper presented at the annual meeting of the nerican Educational Research Association, New York, February, 1971.

easily and elegantly employed to generate testable hypotheses about pupil and teacher behavior. Currently there is a paucity of such models, for they require the careful logical and psychological analysis of complex phenomena. At times the phenomena can only be modeled after lengthy study and integration of prior research. It is this analysis of the problem situation, and the decisions of what to include in a model of observed behavior, what to exclude, and how to juxtapose the remainder, that defines the first conception of multivariate analysis. At least in this sense, multivariate methods are necessary.

The models which have been developed have seen frequent use in educational studies. Among them are the Getzels model of observed social behavior, the Carroll "model of school learning," and the model assumed in Bloom's *Stability and Change in Human Characteristics.*[12] However none contains an analytic description of the facets of the educational environment which impinge upon pupil behavior.

Educational Opportunity

An educational environment exists whenever individuals are present in a setting wherein learning objectives have been formulated and situations developed that are designed to bring about the attainment of the objectives.[13] A large number of settings may be viewed as educational environments—for example the classroom, the home, a museum, and even a peer-group meeting place. One way to view each of these environments is to distinguish two achievement-related components. The first is the simple availability of time, space, materials, and teachers or other behavior models, as the essential resources for learning. The provision of a quantity of high-quality resources to pupils is in one sense *educational opportunity.*

[12] J. W. Getzels. "Conflict and role behavior in the educational setting," W. W. Charters, Jr. & N. L. Gage (Eds.), *Readings in the social psychology of education* (Boston: Allyn and Bacon, 1963); J. B. Carroll. "A model of school learning," *Teachers College Record,* 1963, 64, 723–733; B. S. Bloom. *Stability and change in human characteristics* (New York: John Wiley & Sons, 1964).

[13] J. D. Finn and E. L. Gaier. "The educational environment," Unpublished manuscript State University of New York at Buffalo, 1969.

Educational Press

The second component is the set of activities, expectations, and pressures to which the pupil is subjected, which influence the extent to which he will utilize the resources. This *educational press* is exerted by the learning activities and directions ("read the first ten pages"), general pedagogical characteristics (e.g. degree of organization, enthusiasm of the teacher), as well as by covert psychological demands (e.g. parents' and teachers' expectations for pupil achievement).

This simple model depicts achievement as a function of the factors opportunity and press, i.e.

$$\text{Achievement} = f(\text{Opportunity, Press}).$$

Achievement

The two-component analysis can generate a number of interesting behavioral hypotheses. We might postulate 1) that opportunity is a necessary but not sufficient condition for satisfactory achievement levels to be reached, 2) that press must be neither too low nor too high, for maximal cognitive attainment. Or considering both parameters jointly we may hypothesize 3) that cognitive achievement is maximized when the opportunities provided are at a minimally satisfactory level, *and* the activities and expectations are within a "moderate" range. This is one form of the "consistency hypothesis," the assertion that separate environmental facets must operate in a state of congruence in order for specific outcomes to be realized.

The settings in which a child spends his day may both counteract and support one another. The home, the classroom, and the peer social environment, each provides its own set of resources, as well as activities and expectations for cognitive achievement (Can you name them?) Thus the consistency hypothesis may also be applied across settings; only if the press of the three environments is consistent will learning be expected to be maximized.

Measuring "Opportunity" and "Press"

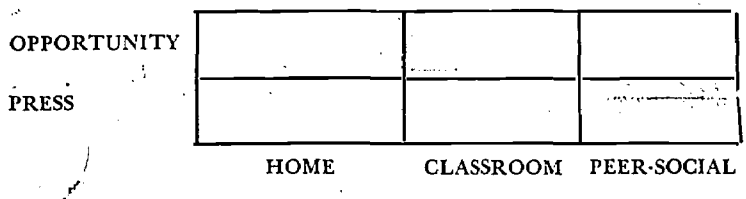
Further definition is necessary to obtain measures of "opportunity" and "press." For example opportunity connotes the availability of human and material resources, as well as a setting with ample time, space, and lighting for children to learn. We may evaluate the instructional materials themselves: are the textbooks, laboratory materials, audiovisual aids of sufficient quantity, of sufficient

quality, and at a level appropriate to the classes in which they are used?

Press is exerted upon a child by the teacher through instruction, and through the manifestation of personal attitudes and expectations. Rosenshine has indicated that clarity of presentation and the degree of teacher enthusiasm are among the most important instructional determinants of pupil performance.[14] Both of these are directly observable. In contrast, teachers may hold subtle expectations for pupils' performance. These shape responses to the class and to individual children, resulting in altered achievement patterns.[15] A less obtrusive approach to the measurement of expectations may be necessary.

The analysis of environmental components may be depicted in block-diagram form.

Figure 1



There exist innumerable research reports involving one or two aspects of the "total picture." The framework can be used as a set of principles by which these are organized. From this, studies may be designed to bring together the effects of multiple environments, or multiple aspects of a single environment. No single study can test any of the broad propositions in its entirety. However, several well-designed studies (with varying definitions of opportunity and press), provide all the necessary information to accomplish this purpose.

Bloom has recommended a similar integrative approach, under a three-part breakdown of the antecedents of learning.[16] He suggests that the maximum proportion of vari-

[14] Rosenshine, *op. cit.*

[15] R. Rosenthal and L. Jacobson. *Pygmalion in the classroom* (New York: Holt, Rinehart, and Winston, 1968); J. D. Finn. "Expectations and the educational environment," *Review of Educational Research*, 1972, in press.

[16] Bloom, "Individual differences." *op. cit.*

ation in pupil achievement determined by pupil entry behavior is 50%; by affective entry characteristics, 25%; and by instructional quality 25%. Study of both entry behavior and affective entry characteristics, indicates that jointly the two may account for 65% of the variation in achievement scores. All three components may account for as much as 90% of the variation in learning outcomes. I personally do not believe that such a high degree of explanation can be achieved without including other school and non-school environmental characteristics (e.g. availability of resources other than the teacher). However the principle remains that we can only achieve an increased level of understanding by developing the means of organize and test more complex behavioral models.

OUTCOMES

Let us turn briefly to the *outcomes* of formal education. For most if not all popular concern with education, involves outcomes which co-vary only partially with basic cognitive achievement. No single outcome is so simple and important as to be worthy of study in a single aspect, or in isolation.

One analysis of educational outcomes is provided by the distinction of cognitive and affective achievements, in the two handbooks of the *Taxonomy of Educational Objectives.*[17] Both define classes of behavior which a pupil *can* demonstrate when placed in an appropriate stimulus situation. Excluded is the entire set of behaviors which the individual actually *does* display, when facing parallel situations in daily life, i.e. "habitual behavior." These too may result from educational experiences, and may develop in specific units or courses, or from longer schooling periods.[18] All three outcomes (cognitive, affective, habitual) may be considered when evaluating a course of instruction or in studying, say, attitudes toward reading or attitudes toward alien cultures.

[17] B. S. Bloom, (Ed.). *Taxonomy of educational objectives. Handbook I: Cognitive domain* (New York: David McKay & Co., 1956); D. R. Krathwohl, B. S. Bloom and B. B. Masia. *Taxonomy of educational objectives. Handbook II: Affective domain* (New York: David McKay & Co., 1964).

[18] J. D. Finn. "Evaluation of instructional outcomes: Extensions to meet current needs," *Curriculum Theory Network,* 1972, in press.

From a different viewpoint, Jackson suggests that the primary outcomes of formal educational settings are adaptive ons; that above all else, the pupil must learn to adapt to the power and organizational structure of a crowded school situation.[19] The associated learnings are some of the "unanticipated achievements" of educational process. Combining these with the taxonomic analysis, the categories of required behavior in school are three: (A) behaviors demonstrating adaptation to the environmental structure, (B) behaviors exhibiting expected performance with the course materials, and (C) behaviors demonstrating particular forms and degrees of interaction with other pupils.[20]

These outcomes are not stressed equally by teachers, by the class as a whole, or by peer groups. For example a teacher may require certain learning criteria to be met, while peer group demands conflict with this. A teacher may provide goal statements in terms of performance with class materials. On the other hand, rewards and punishments may be mediated on the basis of the pupil's accepting the classroom conditions and power structure. It would be useful to examine the press of each constituent upon the child, as well as his responses in the three achievement categories.

In the preceding paragraphs I have provided some examples of one sort of "multivariate analysis," i.e. the logical analysis of the educational setting, and of the outcomes which are realized. Multiple input–multiple outcome models of educational process must be developed, to facilitate relating the two *sets* of constructs, rather than specific facets of either. Propositions derived from such models must be empirically tested as a step toward increasing our understanding of contextual behavior.

THE ANALYSIS OF MULTIVARIATE DATA

Multivariate *statistical* models are applicable to situations with multiple random variables. Most linear models employed to analyze educational data (e.g. regression or analysis of variance models) assume fixed values of the independent (predictor) variables. In contrast multivariate

[19] P. W. Jackson. *Life in classrooms* (New York: Holt, Rinehart and Winston, 1968).

[20] Finn and Gaier, *op. cit.*

models are usually appropriate when a study contains multiple dependent, or outcome, or criterion measures. They frequently constitute the most realistic statistical models for behavioral data, especially when the research evolves from a multiple-outcome paradigm.

Factor Analysis

Statistics journals abound with analysis techniques that are "multivariate." However there are two distinct sets of multivariate procedures currently employed in education and psychological research. The first was developed originally by Spearman and by Thurstone as a procedure for discovering the common behavioral "factors" underlying test responses. Until recently, factor analysis has been employed with multiple outcome measures in the attempt to discover latent score components. The problem is one of identifying some independent variables which are not known or specified a priori. Factor and component analysis techniques have been used as exploratory devices, since they yielded no tests of particular structural models. Instead the researcher would collect some data and use factor analysis methods to "see what comes out." It is possible that Thurstone's efforts were not only among the first, but also one of few productive efforts of that sort in education.

More recently factor-analytic models have become statistical models, as the distributions of the resulting estimates have become known. Psychometricians now have the theory as well as computer programs for hypothesizing particular constructs underlying test responses. With recent advances in maximum-likelihood factor analysis and covariance structure analysis,[21] hypotheses may be tested about the existence and form of the underlying factors, and about the congruence of several data factorizations. Factor analysis can now assume a more useful and revealing role in attempting to describe response data.

[21] K. G. Joreskog. "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika,* 1969, 34, 183–202; R. P. MacDonald. "Numerical methods for polynomial models in nonlinear factor analysis," *Psychometrika,* 1967, 32, 77–112; R. D. Bock and R. E. Bargmann. "Analysis of covariance structures," *Psychometrika,* 1966, 31, 507–534.

ANALYSIS OF VARIANCE AND REGRESSION

The second set of multivariate models includes generalizations of analysis-of-variance and regression procedures to the case of multiple dependent variables. For example, with several groups of subjects we may wish to test that the group means are equal, not on a single outcome variable, but on two or more intercorrelated variables simultaneously. For this a multivariate extension of the usual F-test may be employed. Or we may wish to predict achievement in a literature course from one or more instructional variables. Achievement may be measured by both an indicator of cognitive performance, as well as an index of the pupils' attitudes toward reading literary material. For this a multivariate extension (here bivariate) of the multiple linear regression model is appropriate. We may test for the simultaneous prediction of the two (interdependent) outcomes, which jointly describe the results of the course or unit.

Multivariate data arise whenever more than a single measure, test, or subtest, is necessary to describe the outcome of an experiment or comparative study. This may occur in several ways. First, multiple measures are taken when there is uncertainty about the most appropriate operational definition of a construct. For example, in the absence of a single good definition of general intelligence we may administer several cognitive tests, each having a heavy "g" component. Second, multiple scales may be necessary to complete the description of responses to a given stimulus. For example, multiple measures derived from the Rorschach or from unstructured language samples[22] may together yield a conception of underlying behavior traits. Third, multivariate data occur when there is practical interest in more than one distinct outcome of a study. This situation is encountered frequently in educational studies. Outcome measures may consist of both cognitive and affective scales; of both higher and lower mental process cognitive measures; of differing aspects of reading skill; or of both anticipated and unanticipated instructional outcomes. Fourth, problems of nosology, encountered in grouping pupils, are multivariate problems when multiple classificatory variables are considered.

[22] J. D. Finn. "Patterns in children's language," *The School Review*, 1969, 77, 108 127.

Finally "repeated-measures" research by definition yields data which are multivariate. Repeated measures problems arise whenever subjects are measured more than once on the same scale. For example, the same instrument may be administered as pre- and post-test for a course or unit. Learning studies frequently allow the subject multiple trials on the same task, with the same measure of proficiency at each trial. Most long-term and longitudinal studies involve administering the same test or set of stimuli to the same subjects, repeatedly. The repeated-measures problem is like other multivariate situations, except that the multiple outcome variables are measurements on the same scale.

In each example the same observational units are measured on more than a single test, or sub-test, or occasion. It is this property that defines the problem as multivariate in the statistical sense. A multivariate approach attends to the data as a whole, rather than to a few isolated or transient aspects. In each instance the analysis of a single summary measure, i.e. a total or average score, will result in the loss of information conveyed by the individual scales. Analyses of each of a series of measures separately, will result in redundancy which in turn will threaten the validity of the conclusions drawn. The appropriate multivariate model will allow the researcher to retain the multiple scores, and to treat them simultaneously. There is no particular restriction upon the intercorrelations of the measures. Since multivariate statistical outcomes pertain to the set of variables, it is only important that the set be a meaningful one.

Multivariate Models

The particular multivariate models considered here are extensions of the univariate analysis-of-variance, covariance, and linear regression models. There is often some confusion of terms in referring to these models, and I would suggest the following conventions. The number of dependent variables (which are random variables, or "variates") determines whether a model is univariate (1) bivariate (2) or multivariate (generally). In regression the number of predictors determines whether the model is simple regression (1), or multiple regression (generally). Thus the regression model with one predictor and one outcome is the univariate simple regression model; with one predictor and

multiple outcomes, the multivariate simple regression model; with many predictors and many outcomes, the multivariate multiple regression model; etc. The analysis-of-variance counterpart of univariate simple regression is one-way univariate analysis of variance; a one-factor design with many criterion measures requires a one-way multivariate analysis of variance model; a two-way or many-way design with multiple outcome measures from each subject, requires a many-way multivariate analysis of variance model; and so on.

There is a direct parallel between the regression and analysis-of-variance models. The only distinction between them is a conceptual one. In regression analysis the independent variables (predictors) are measured variables; in analysis of variance the independent variables have only discrete group-membership values. Algebraically, the two models are the same, and may be solved by a single estimation algorithm. The F-statistics which usually result from the two analyses are mathematically identical, even though the associated terminology sometimes differs.

In multiple regression there is no restriction upon the intercorrelations of the predictor variables. Also in analysis of variance (or covariance) the independent variables may be correlated. This means that (contrary to the presentations of many statistics texts) valid and exact analyses of variance may be conducted with *unequal* subclass frequencies, and with contrast parameters which are not orthogonal. The linear statistical model with one or more intercorrelated independent variables of either sort, and one or more intercorrelated criteria is the "multivariate general linear model." All analysis-of-variance, covariance, and regression problems may be solved as special cases.

The results of a multivariate statistical analysis are essentially the same as those from the usual univariate techniques, but apply to multiple outcome measures. For example, in a 2 × 3 (sex × social class) experimental design, with three outcome measures from each subject, univariate estimates of the sex, social class, and interaction effects may be obtained. The hypotheses may be tested that μ males $= \mu$ females, that means for the three social class groups are equal, and that the interaction terms are null. The multivariate model will yield a single F-ratio for each of these hypotheses, as a summary test statistic apply-

ing to all three dependent variables. In addition, good general-model programs for multivariate analysis yield the usual univariate F-statistics for each of the measures, to facilitate interpretation of the multivariate result.[23] The univariate ratios however, lack statistical independence to the extent that the variates are intercorrelated. By comparison the multivariate test statistics are not invalidated by non-zero intercorrelations of the dependent variables. These correlations would in fact be expected, if the multiple measures are facets of a common trait or response.

Multivariate test statistics may also be obtained for a single planned comparison among groups, for the regression of multiple criteria upon one or more predictor variables, and for mean differences on multiple criteria after removing the effects of one or more measured control variables (covariates). In every case, when the number of dependent variables is one, the multivariate results reduce to the usual univariate form in both theory and computation.

CANONICAL ANALYSES

There are some multivariate techniques in current vogue which seem to be employed more than their power would warrant. These are generally referred to as canonical analyses, of which two are canonical correlation analysis and discriminant function analysis. Both of these are procedures for identifying linear combinations of criterion measures which have specified optimum properties.

Canonical Correlation

When the research yields both multiple dependent variables and multiple measured predictors (the multivariate multiple regression), the canonical correlations are measures of association of the two sets. Linear functions of the criterion variables are identified, and other linear functions of the predictors, such that the two resulting functions are maximally intercorrelated. For example, if the criterion variables are $y_1, y_2, \ldots, y_p$, and the predictors are x_1,

[23] J. D. Finn. *"Multivariance—univariate and multivariate analysis of variance and covariance: a Fortran IV program. Version 5,* (Ann Arbor Michigan: National Educational Resources, Inc., 215 Kenwood Ave., March, 1972); D. J. Clyde, E. M. Cramer and R. J. Sherin. *Multivariate statistical programs* (Coral Gables, Florida: University of Miami, Biometric Laboratory, 1966).

$x_2, \ldots, x_q$, new variables Y and X are defined, such that

$$Y = a_1y_1 + a_2y_2 + \ldots + a_py_p$$

and

$$X = b_1x_1 + b_2x_2 + \ldots + b_qx_q.$$

The weights a_j and b_j are computed in such a way as to maximize r_{XY}, the simple Pearsonian correlation of X and Y. r_{XY} is a canonical correlation between the two sets of variables. The significance of r_{XY} may be tested in the usual manner, to determine whether there is some non-zero association of the two sets. If there is only a single y-variate (univariate case), the canonical correlation is simply the multiple correlation coefficient, and the b-weights are the partial regression coefficients. Further if there is also only a single x-variate, the canonical correlation is the simple Pearson product-moment correlation of x and y.

Although canonical correlations are generalizations of more common correlation measures, their interpretation in the multivariate case is difficult. To completely describe the association of the two sets of variables, as many different linear functions and correlations are required as the number of variables in the smaller of the two sets, i.e. min (p,q). More than one of these, and perhaps all, may be significantly different from zero. For each significant correlation, the function and its weights require interpretation. The weights are computed simultaneously and reflect both the scale of the original measures, as well as the particular set of measures taken and their intercorrelations. The magnitudes of the weights do *not* generally reflect the relative contributions of the measures to r_{XY}. Interpretation of the weighted linear function is subject to all the ambiguities of the naming of factors in factor analysis.

A more revealing approach is to obtain multivariate test statistics for the regression of the criterion measures upon the predictors. The simpler univariate estimates and test statistics should provide a clear interpretation. These may include simple, partial, and multiple correlations, percentages of variation accounted for, and so on. While sophisticated procedures such as canonical correlation analysis may appear elegant, our purposes are best served by the *simplest* and *most straightforward* techniques which yield valid answers to our research questions.

Discriminant Analysis

Discriminant analysis bears general similarities to canonical correlation analysis. When the independent variables have group-membership values, as in analysis-of-variance designs, linear combinations of the criterion measures may be obtained which themselves have maximum between-group variation or maximum univariate F-statistics. Again multiple "discriminant functions" may be necessary to explain all between-group variation in a set of measures. Each is subject to the interpretive complexities of canonical correlation weights. The test statistic for the significance of all discriminant functions is identical to the ordinary multivariate test of the equality of mean vectors. No between-group variation is created or destroyed through the transformation to discriminant variables.

For interpretive purposes in multivariate analysis of variance, the magnitudes of multiple univariate F-ratios may be compared. In this way the variates most and least effected by the group-membership variable can be identified. Linear combinations of means, or *contrasts*, provide the researcher with estimates of the strength and direction of between-group effects. The corresponding standard errors may be employed to provide additional information about the strength of effect relative to replication of the study. Strength-of-association measures such as ω^2 are not generally valid in fixed-effects anova.

A good introduction to multivariate analysis, with examples, is given in Bock and Haggard.[24] Studies using multivariate methods include those reported by Finn, and Harrison.[25] Texts are appearing with increasing rapidity.[26]

[24] R. D. Bock and E. A. Haggard. "The use of multivariate analysis of variance in behavioral research," D. K. Whitla (Ed.). *Handbook of measurement and assessment in behavioral sciences* (Reading, Mass.: Addison-Wesley, 1968).

[25] Finn. "Expectations,"; F. I. Harrison. "Aspirations as related to school performance and socio-economic status." *Sociometry*, 1969, 32, 70–79.

[26] R. D. Bock. *Multivariate statistical methods in behavioral research* (New York: McGraw-Hill, in preparation), Chapter 7; J. D. Finn. *The analysis of multivariate data* (New York: Holt, Rinehart and Winston, 1974, in preparation; S. J. Press. *Applied multivariate analysis* (New York: Holt, Rinehart and Winston, 1972); M. M. Tatsuoka. *Multivariate analysis: Techniques for educational and psychological research* (New York: John Wiley and Sons, 1971).

APPLICATIONS

Only after we have chosen a conceptual approach to a problem are we ready to decide upon a specific analysis mode. The general linear model may be employed whenever there are multiple antecedent and/or multiple outcome variables. With a good general-model computer program, it is not necessary to score and code data to fit software specifications. Instead the program may be adapted to the data as they are found. When a linear-model analysis is not appropriate, non-linear models, non-parametric models, time series models, and others may be required. Our purpose is to describe the trends which exist in quantitative data. While a statistical model must be selected according to the design of a study, *any* valid procedure for discovering relationships which exist, may be useful. Further, the simplest graph or descriptive device should be used to communicate trends which may have taken a researcher many complex analyses to discover!

In planning research, a number of technical issues arise which must be resolved before a study can be completed. For some of the issues the arguments may seem unclear, while for others traditional procedures may seem inappropriate. Some of them have solutions.

Hypotheses

Researchers are often able to design studies elegantly but will assert that they have no idea what effects to expect, i.e. no hypotheses. If one discusses the study with them informally, they may admit that they really believe the new approach to be better than the old, or that their results will be essentially the same as another investigator's in a different situation. These are hypotheses. Hypotheses are informed best guesses as to the anticipated experimental outcomes. Frequently a problem yields competing hypotheses, each of which would suggest a different outcome. They too should be stated and tested as competing explanations.[27] By so doing, both statistical and logical confirmation power are preserved. To confirm a hypothesis is to twice support an assumption, once logically and once empirically. And nothing is lost; no amount of exploration or estimation is precluded by testing specific prior beliefs.

[27] Platt, *op. cit.*

Comparing Instructional Approaches

Instructional approaches with specific goals must be examined by goal-specific analysis techniques. For example, a "mastery-learning" approach is effective if variation among pupils is decreased, with no sacrifice of mean achievement level. To evaluate a mastery approach, within-class *variances* must be estimated and compared. Many instructional approaches provide effective experiences for all children regardless of ability or intelligence. These approaches are effective if they yield lower *correlations* of ability and learning, than comparison methods. Here correlations must be compared across treatments. In both examples, contrasting means across instructional approaches is only a partial and secondary analysis.

Much attention has been given to the measurement of change.[28] Individual gain scores are notoriously unreliable. When focus is upon individual achievement, "true gain scores" should be estimated.[29] However *mean* gain scores for a group of pupils are likely to be much more reliable. Since mean change is basic to much educational process, simple and regressed change scores should be employed whenever they logically reflect the criterion of concern.

Comparative vs. experimental studies

Experimental studies or "true experiments" are conducted when the independent variables are manipulated and assigned to subjects by the researcher.[30] *Comparative studies* involve the correlation of measures, or comparison of groups which exist independently of the experimental situation i.e. the comparison of intact groups). Only in an experimental situation can the outcomes be directly attributed to the manipulated antecedents. Both approaches are needed in education.

There are frequent calls for more experimentation;[31] yet there are situations in which experiments raise critical

[28] L. J. Cronbach and L. Furby. "How we should measure "change"—or should we?" *Psychological Bulletin*, 1970, 74, 68–80; Bock, *op. cit.*; E. F. O'Connor. "Extending classical test theory to the measurement of change," *Review of Educational Research*, 1972, 42, 73–97.

[29] L. R. Tucker, F. Damarin and S. Messick. "A base-free measure of change," *Psychometrika*, 1966, 31, 457–473.

[30] D. T. Campbell and J. C. Stanley. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.

[31] J. C. Stanley. "Quasi-experimentation in educational settings," *The School Review*, 1967, 75, 343–352.

questions that can only be resolved through the comparative approach. For example, Rosenthal and Jacobson experimentally manipulated teachers' expectations for pupil performance.[32] The results observed were gains in both pupil achievement and ability levels. Much skepticism was evoked among researchers, that so small and experimental treatment could have so large an effect. Teachers' expectations are reinforced on a daily basis by observations and interactions with the pupils. These experiences would tend to counteract a small-scale artificial manipulation. Thus teachers' expectations and their effects can only be studied in their natural setting and monitored regularly for their interactions with pupil performance.

The mode of statistical analysis does not define whether a study is experimental or comparative, nor does it determine whether cause-effect relationships may be established. Analysis-of-variance techniques may be employed whenever group means are to be compared, whether the groups are naturally or experimentally constituted. Correlations may in fact represent causation, if the independent variable is manipulated and assigned to subjects by the experimenter.

Measurement scales

When measured data are collected they should not be dichotomized, trichotomized or otherwise partitioned into a smaller number of score categories. For example, if measured scores range from one hundred down to zero, (100 99 ... 68 67 | 66 65 ... 35 34 | 33 32 ... 2 1 0), we might be inclined to split into the highs, lows, and the remainder, at scores of 33 to 66. To do so is to treat a score of 67 as essentially the same as a score of 100, but as different from 66. With statistical tools such as the multivariate general linear model, all measures may be analyzed in the scale in which they occur.[33]

Unit of analysis

When an instructional approach is applied to an entire class or any defined group of pupils under one set of conditions, the class or group is the analysis unit, not the indi-

[32] *Op. cit.*

[33] An exception to this rule may arise in the study of extreme groups; c.f., F. I. Harrison. "Opportunity as it is related to home background and school performance," *The School Review*, 1969, 77, 144–151.

vidual pupil.[34] In group settings pupils are not independent observations. If one child raises a particularly cogent point, most or all of the pupils reap the benefits. If the teacher is absent for one lesson, all the pupils suffer. To have only a single experimental group threatens the external validity of the study. Multiple groups, meeting at various times of day, and each with a different teacher, are necessary. Either the class mean or standard deviation, or means of subgroups within a class, are the appropriate analysis units. Each class contributes only a small number of independent observations to the analysis.

Selection of a method

It is deceptively simple to become method-oriented. "We could do a case study," "I ran the XYZ computer program, but it didn't tell me what I wanted to know," "Statistics don't tell us anything about education." These statements and many more like them reflect an important logical reversal. *All* of our methods of research and analysis are tools which can be applied when needed. Perhaps the best educational program for our students is exposure to the maximum *variety* of research techniques. More important, however, is continued experience in the analysis of educational problems! Each must be evaluated in deciding what techniques may be useful, or indeed whether and how empirical research can contribute to its solution.

SUMMARY COMMENTS

In the first sections of this paper I discussed some directions in which I feel we should move, in research and evaluation in the language areas. I have considered "evaluation" in a broad sense, and much of the discussion is on a general level. I hope that the ideas expressed are not so complex that one cannot give them serious consideration in one's own research activities. I feel very strongly that they are necessary directions in assessing both the scientific significance as well as the "relevance" of our work. I would suggest for example, that the general paradigm presented

[34] J. Raths, "The appropriate experimental unit," *Educational Leadership*, December, 1967, 263–266; Stanley, *op. cit.*; D. E. Wiley and R. D. Bock, "Quasi-experimentation in educational settings: Comment," *The School Review*, 1967, 75, 353–366.

does not describe a theoretical and undefinable situation. Instead its components can be seen and measured in many forms, in the inner-city school, in the open-plan classroom, and indeed in a child's own bedroom.

I have chosen not to discuss the technical issues of measurement. It is certainly an important topic, expecially when we attempt to gather data on multiple aspects of the educational environment. I refer readers instead to an outstanding new book which depicts many of the measurement alternatives that are being employed.[35] I assume that you will be sufficiently innovative to gather multiple reliable measures of the objects and events of your concern. Instead I have attempted to describe some of the newer techniques for data analysis. These are particularly relevant to the analysis of data collected under a multiple input-multiple response model of education.[36]

[35] R. M. Brandt. *Studying behavior in natural settings* (New York: Holt, Rinehart and Winston, 1972).

[36] I am grateful to Professors Paul Lohnes and Forest Harrison for their helpful comments.

Research into Imagic Association And Cognitive Interpretation

HARRY S. BROUDY
University of Illinois, Urbana

One of the benefits expected from educational research is unimpeachable evidence for or against the usefulness of this or that school practice. Research helped to eliminate the study of Latin from the high school, disclosed that the average man uses only about 10,000 words in his correspondence, makes only certain kinds of grammatical and spelling errors, does only certain kinds of arithmetic problems. Since the cult of efficiency has never lacked for devotees in American education, this kind of research was praised for removing alleged deadwood from the curriculum and bringing schools to that happy state in which only that would be studied which has a very good chance of being used in life. To cap it all, Edward M. Thorndike's theory of transfer by identical elements made it plausible to practice only those processes which would be replicated in nonschool tasks. For Thorndike, the best transfer was no transfer.

THE PREDICAMENT OF GENERAL EDUCATION

However, this plausible doctrine and no-nonsense strategy have never convinced the proponents of general education or, for that matter, the advocates of the theoretical foundations of professional training. Nonetheless, the school was jockeyed into the awkward position of having to demonstrate that something learned formally in school (input) functions in life or the nonschool situation (output). The current to-do about overqualification of people for jobs and the high rate of unemployment among recent

college graduates has rendered the need for such a demonstration critical. Indeed, colleges have already responded by trying to make school as much like life (relevant) as possible by making the life task itself the subject of instruction. This strategy merely increases the skepticism about the need of higher education. CBS last May presented an hour program titled: "Higher Education: Who Needs It?" The answer the viewer was led to infer was: "Not many."

Clearly the rationale for the academic enterprise rests on C. H. Judd's rather than Thorndike's theory of transfer. Judd argued that transfer occurs when a given situation is interpreted in the light of some theory or principle from which it is directly or indirectly deducible. Hence if one learns the right principles, he will be able to transfer their import to a wide variety of life situations. On this view, general or liberal education is highly generalizable schooling. However, if the particular situation or problem or task can be solved by a procedure that can be imitated or performed according to a rule, knowledge of the generalization is unnecessary. Thus one need not know the theory of oxidation to put out a fire with a chemical extinguisher; one merely needs practice with a fire extinguisher.

The current enthusiasm for performance-based teacher education, the detheorization of professional curricula in many fields, the revolt against the honored place of theory and the kind of minds at home with theory are all bolstered by, if not founded on, this contingent relation between the doing of X and knowing why one does X in a given way. Since knowing how to do X covers the standard cases, society can get along with small cadres of professionals who know *why* (for the nonstandard cases) and large contingents of paraprofessionals and technicians (practice people) who are guided by rules and technologies devised for them by the theory people.

Where does this line of argument leave general education and the language arts as part of general education? General education seems to be a needless luxury, and the language arts must be trimmed to what the average man uses in his average communication behavior. No doubt research would show us that for ordinary discourse the

level of language usage need not be high by any standard. Popular language has ready-made phrases wherewith to express ready-made responses; slang provides spice, and the electronic visual media provide the picture that surpasses a thousand words. Why people think that it is important for Johnny to read well, on this line of reasoning, is a mystery—unless, of course, the reasoning is faulty.

"STANDARD ENGLISH"

After examining the language uses by educated and noneducated groups, Fries came to several striking conclusions more than 30 years ago:

1. He rejected the notion of a "correct" English usage and substituted for it ideal "standard English." He explained that standard English historically is a "local [London] dialect, which was used to carry on the major affairs of English life and which gained thereby social prestige."[1] In our own country standard English is used in the conduct of the important affairs of our people, but Fries adds: "It is also the type of English used by the *socially* acceptable of most of our communities and insofar as that is true, it has become a social or class dialect in the United States."[2]

2. Fries notes:

> In the matter of the English language it is clear that anyone who cannot use the language habits in which the major affairs of the country are conducted, the language habits of the socially acceptable of most of our communities, would have a serious handicap . . . The schools, therefore, have assumed the burden of training every boy and girl, no matter what his original social background and native speech, to use "standard English," this particular social or class dialect.

Some people, according to Fries, believe this to be an impossible task and "the almost unanimous condemnation of the results of their efforts convinces us that either the schools have not conceived their task adequately or they have chosen the wrong materials to accomplish it."[3]

---

[1] C. Fries. *American English Grammar* (New York: D. Appleton Century, 1940, p. 12).

[2] *Ibid.*, p. 13.

[3] *Ibid.*, p. 14.

3. In its materials, also, this study seems very clearly to show the point of primary attack in a program of English language for the schools should be in striving to develop a knowledge of practice in using the *wide resources of the language*. Over and over again in the preceding chapters it appeared that the differences between the language of the educated and that of those with little education did not lie primarily in the fact that the former used one set of forms and the latter an entirely different set. In fact, in most cases, the actual deviation of the language of the uneducated from Standard English grammar seemed much less than is usually assumed, and in practically all instances was in the direction of greater conservatism. Vulgar English uses many forms that were common in the older stages of the language and that Standard English has given up.[4]

4. The most striking difference between the language of the two groups lay in the fact that Vulgar English seems essentially poverty stricken. It uses less of the resources of the language, and a few forms are used very frequently. Get, for example, in its many senses appears in both Standard and the Vulgar English materials, but it is employed ten times as frequently in the Vulgar English letters as in those of Standard English . . . In vocabulary and in grammar the mark of the language of the uneducated is its poverty. The user of Vulgar English seems less sensitive in his impressions, less keen in his realizations, and more incomplete in his representations. It would seem to be sound inference from the results of our study that perhaps the major emphasis in a program of language study that is to be effective should be in providing a language experience that is directed toward acquaintance with and practice in the rich and varied resources of the language. Here again observation of actual usage seems the most important method, and the tools of observation absolutely essential.[5]

Granting that standard English is the language used by a social class in our society, does it follow, as Fries concludes, that the public schools ought to teach it to every boy and girl, regardless of his or her social origin? And if this goal has been accepted, why does Fries say that the schools have been remarkably unsuccessful in this effort?

[4] *Ibid.*, pp. 287–288.

[5] *Ibid.*, p. 288.

Non-researchable and Researchable Questions

The first question does not lend itself to research inasmuch as it involves complex value issues. About the only empirical data that are relevant to it are the *opinions* about whether one should "force" standard English on those to whom it is almost like a foreign language, or at least a strange dialect of their native tongue. Some militant blacks, for example, see no compelling reason for black children to learn standard English; the ghetto version of English is richly expressive in its own right. If this ghetto dialect works to the detriment of these children competing in a middle-class society, the militants reply: "Let's change the society." If the data bear out these conjectures, then Fries' goal of language instruction—the public school's goals, in fact—may be up for revision, because the arguments that Fries thought persuasive would no longer hold.

The second question does invite research. Might it be the case that schooling has less effect on language usage than the home or the peer group or the general environment? Might it be the case that only if work and nonschool life reinforce standard English does it become the modal usage of the individual? Does the child of poverty speak one form of English in the classroom and another on the playground or at home? To what extent does the language dialect used on television affect usage in adult or late adolescent life? Most important of all, does mastery of standard English improve communication or change the quality of individual life?

If the choice between standard English and other dialects is merely a matter of cultural convention and does not really affect the quality of life of the user, then the circumstance that the important affairs of the community are couched in standard English is of secondary importance. If the difference is merely that of linguistic style, there can be translation of important events in many linguistic media. The case would be otherwise, however, if linguistic usage were somehow tied up with the cognitive repertoire of the speaker or reader. "Gross National Product," for example, is a concept that compresses a complex of other concepts. Does it have a counterpart in uneducated language use? The difficulty with nonstandard English is not that it lacks the linguistic apparatus for naming

concepts, but rather that it may lack the concepts themselves. If the nonstandard dialects are characteristic of subcultures that do not form the concepts in which the important affairs of the culture are couched, then we are arguing about something more than names. Can we test this hypothesis? Does the correct use of certain concepts in science, economics, history, literature require the use of standard English (or the educated uses of any other language, for that matter) or does the pupil translate these concepts into his own dialect without loss of significance or precision? The defense of the educated use of language—and therewith of general education—depends on the answers to such questions.

Linguistic Poverty

Let us turn now to another finding of the Fries study quoted above. He remarks that the difference between the educated and vulgar use of the language is not so much a difference of grammatical precision or communicative efficacy as of richness or poverty of linguistic resources. But what is the value of richness if it does not necessarily improve communication? I have already suggested that if the poverty of language is accompanied by a poverty of concept the communicative efficiency may be impaired. But there is another consequence of linguistic poverty, viz., that it may denote a poverty of associative resources and therewith a diminished power to generate meaningful connotations in response to linguistic signals. Thus the reading of literature in general and poetry in particular would be directly affected by linguistic poverty, but the cognitive uses of language are also affected by the volume of what Herbart called "the apperceptive mass." Both art and science depend on the imaginative powers of the mind, and although language is unable to label all that the imagination can conjure up, it is by far the best catalogue of experience we have. Poverty of linguistic resources may betoken a poverty of thought and feeling as well.

But can research support our intuitions in these matters? Yes, if it turns its attention to the uses of schooling in nonschool situations and especially in the adult use of schooling. But unlike the surveys of the Twenties that sought

to make average usage a norm for instruction, we need to bring out the differences between the educated and the uneducated uses of schooling. We cannot conscientiously urge that all of our young people exert the effort to become educated, if it makes little or no difference in the quality of their thought and feeling in the vocational, civic, and personal dimensions of their experience.

USES OF SCHOOLING

Replicative and Applicative Uses

Much of the research on the uses of schooling in nonschool life has been vitiated by the supposition that a school input (language or history or mathematics) is used (a) by reinstating on cue the elements of what has been learned. This can be called the replicative use of schooling. Or (b) by applying some principle to solve a problem. It is assumed that if the study of mathematics in school was successful, the person can survey a plot of land or calculate his income tax correctly or do statistical analysis.

I have argued elsewhere[6] that schooling as used in life (apart from our vocational specialties) for the most part is neither replicative nor applicative. Given a nonschool task T, we do not, as a rule, reinstate a school learning L which is a part of T, nor do we reinstate a generalization G from which we deduce a solution to T. If replication is used as a criterion for the efficacy of schooling, then most of schooling was a waste of time because tests for retention of specific items of school learnings give disappointing results. If application of theory is a criterion, schooling is also a waste of time, for only when we are very familiar with a domain of phenomena and only when we possess the technical skill to deal with them can we be said to be applying knowledge.

Associative and Interpretive Uses

I would argue—and I would hope research will confirm—that as far as general education is concerned, two other uses of schooling are the functional ones: the associative

[6] H. S. Broudy, B. O. Smith, and J. R. Burnett. *Democracy and Excellence in Education* (Chicago: Rand McNally, 1964); H. S. Broudy. "The life uses of schooling as a field for research," *Philosophical Redirections of Educational Research, 71st Yearbook, N.S.S.E. Pt. I.* 1972, 219–239.

and the interpretive. Briefly, the associative use of schooling reinstates in nonschool situations those meanings, images, and words that for one reason or another can be associated with the stimulus. The interpretive use of schooling supplies the appropriate context and concepts for categorizing and analyzing a given nonschool task. If empirical research supports this hypothesis, then we might begin to fashion a more plausible rationale for general education, because we could then claim that until these (associative and interpretive) uses of schooling are taken adequately into account, it is at least premature to dismiss general education as useless. The dogma of behavioral objectives sees to it that they are not taken adequately into account.[7]

There are two points to be argued here. One is that the associative and interpretive uses do in fact occur; another is that they are important. Let me comment on the second point briefly. The quality of response, and indeed of all individual life, can be assayed roughly by the richness of the imaginative perception that goes into it, on the one hand, and by the relevance and cogency of reasoning on the other. As to why opulence of imagination and cogency of thought are valuable, one would think not much would have to be said, but in view of the virulence of the attack on general education, some explicit defense is in order. The argument is simply that in a technologically dominated society about the only avenues to sanity, not to speak of individuality and authenticity, are (a) new forms of behavior suggested by creative imagination, and (b) commitment based on rational appraisal of the forms so suggested.[8] If this is what general education promises—and no other sort of schooling does promise it—then the school will have justified its mission, if research shows that in fact the input is so used.

[7] H. S. Broudy, "Can research escape the dogma of behavioral objectives?" *School Review*, 1970, 20, 43–55.

[8] This is not the place to elaborate this thesis, but by new forms I mean, for example, the behavioral styles that the new woman may adopt or the behavioral forms that will come to be regarded as appropriate for the enactment of such virtues as chastity, courage, temperance, patriotism, and the like.

USES OF LANGUAGE ARTS SCHOOLING

So much then for general considerations. As for the curriculum in language arts and literature, perhaps it would be advisable to examine the inputs in terms of the replicative, associative, interpretive, and applicational uses one might expect the adult to make of them. This is of some importance, because items that are to be used pretty much as learned, spelling, for example, have to be overlearned. This would be true also of some other language skills, but how many of these are there, and to what level of retention do we need to develop them? Granted that one does not learn to write well by memorizing the rules of grammar, does it follow that being able to replicate some rules might not help? To take an example from another field: memorizing the rule for the expansion of a binomial does not help the student to understand why the rule works, but insofar as computation is concerned, the rule helps enormously, even if learned and used by rote. Rote learning and the drill that produces overlearning are not regarded as high-grade intellectual functions, and it is fashionable to denigrate them, but as every specialist in a field knows, once the rationale behind a rule is understood, it is not necessary or economical to repeat the demonstration every time one needs to use the rule. It would help considerably for research to arrive at some reliable estimate as to just how much of this overlearning of rules, forms, and other constants in language use it is profitable to supply. This knowledge would permit us to avoid the Scylla of relying on rote memory altogether and the Charybdis of hoping the student will intuit rules from his own trials and errors.

There are also some facts about literature that are used replicatively. Although the names of works, authors, places, the classification of styles and periods are not on the same plane as appreciation of literary works, they cannot be dispensed with entirely. The question is not so much whether or not some of these facts *about* literature should be taught for replication, but rather how many and for what length of retention. Times change rapidly enough so that it is not amiss to check periodically the replicative use of such items by the educated adult population.

The study of formal grammar is a fair example of knowledge that is usually justified by its applicative use. If one knows the rule or principle of declension, for example, one will apply it to the proper use of nominative and accusative or dative forms of nouns and pronouns. The logic of grammar would enable the person to test for the appropriateness of dependent clauses, references of pronouns, and the like.[9]

A SMALL SURVEY OF RESEARCH

By and large, however, one has to make the case for the study of composition and literature on the associative and interpretive uses. We need research to tell us how work in these areas affects the response of the adult in reading, discussing, and thinking about what he reads and discusses.

Yet in scanning Henry C. Meckel's chapter on research on teaching composition in *Handbook of Research on Teaching*,[10] I find the following section headings: Methods of Teaching Composition, Interrelationship Among Language Skills, Composition Skill and Maturity, The Relation of Grammar to Composition Skill, with the following subheadings: knowledge of grammatical terminology and ability to parse, the utility of diagramming, grammar and punctuation, determination of critical items of usage, faults in sentence structure, the efficiency of methods, the transfer value of grammar. Other section headings include The Relation of Structure to the Expression of Ideas, The Effects of Practice, The Composition Assignment, Marking and Evaluating Student Writing, Preparation of Composition Teachers.

I do not know how representative this list is of research on the teaching of composition, but it seems not to stress the associative or the interpretive uses of the language arts in nonschool situations, although the transfer value of grammar might be an exception.

Under the general rubric of research on the teaching of

[9] Apropos of which I can report that the most discernible effect on a group of college girls of a course in formal logic that I once gave was reported by their instructor in English composition. Presumably, part of poor writing may be due to poor thinking.

[10] N. L. Gage (ed.). *Handbook of Research on Teaching* (Chicago, Rand McNally, 1963).

literature we find: The Objectives of Instruction, Reading Interest Studies, Extensive and Intensive Teaching of Literature, Surveys of General Classroom Practice, Responses to Literature, The Discussion Method, Factors in Comprehension of Literature, and Promising Research Areas.

As far as objectives are concerned, the associative and interpretive uses are uppermost. Thus in one reorganization of the secondary English curriculum the objectives included the intent to

> ". . . broaden, deepen, and enrich the imaginative life of the student"
>
> ". . . raise the level of appreciation and enjoyment in reading"
>
> ". . . provide intellectual and spiritual stimulation"
>
> ". . . encourage the development of intellectual faculties: sensitivity, imagination, thinking, and interpretation."[11]

The nearest approach to my concern was in the section called Response to Literature. But here too the data were largely gathered through questionnaires given to students,[12] and although some generalizations regarding the response were attempted, for the most part the investigators were overwhelmed by the number of variables involved and the variability of the responses. Perhaps the study cited which is closest to the approach I have in mind is I. A. Richard's *Practical Criticism: A Study in Literary Judgment* (1930) in which he examined the interpretation of poetry by Cambridge students. Richards calls attention to the "misleading effects of the reader's memories of some personal scene or adventure; of erratic associations and stereotyped responses; of confusion caused by doctrinal dispositions of the reader, especially when his beliefs conflicted with those presented in a poem; and of the

[11] J. F. Hosic (Compiler). *Reorganization of English in secondary schools.* (Washington, D.C.: U. S. Government Printing Office, Bureau of Education Bulletin, 1917, No. 2).

[12] Lou LaBrant's survey of the effect of teaching literature on the life habits of students is an exception. That study showed that the method used (highly individualized reading) had a lasting effect on reading habits. L. L. LaBrant. "The use of communication media," In M. Willis, *The Guinea Pigs after 20 Years* (Columbus: Ohio State Univ. Press, 1961), Ch. 8.

effect of general critical misconceptions and technical judgments."[13]

Research Into Adult Uses of School Learnings

Research into the adult uses of school learnings in the language arts is a highly complex affair, and I doubt that we are ready to do sophisticated studies with highly controlled samples on single variables. The research I have in mind is an exploratory approach to more refined studies. As a rough working hypothesis I would argue that for any molar problem, i.e., for any societal or existential problem, to become intelligible, the reader must construct contexts that are appropriate to it. Thus the problem of environmental pollution can be discussed in the political, technical, economic, and aesthetic moral contexts. Linguistic materials on these problems presuppose that the reader can supply the categorial and evaluative schema; hence the author alludes to these schema without making them explicit. Can the reader fulfill these expectations?

I would, therefore, begin with linguistic materials at various levels of sophistication: the ordinary daily newspaper, news magazines, the *New York Times,* scholarly magazines, books, etc. Subjects with varying amounts of formal schooling might be asked to read and respond to these materials with a view to finding out the associative and interpretive resources they bring to these materials. We might, if lucky, discover some clues for relating the responses to the amount and kind of schooling, and it may well be that the blocks to understanding will tell us as much as the successful trials.

As rough illustrations of what I have in mind, consider the following reading task:

> Charles Lyell frequently explained natural phenomena as due to "creation" and, of course, a carefully thought-out creation. The fact that the brain of the human embryo successively passes through stages resembling the brains of fish, reptile, and lower mammal discloses in a "highly interesting manner, the unity of the plan that runs through the organization of the whole series of vertebrated animals; but they lend no support whatever to the notion of a gradual transmutation of one species

[13] I. A. Richards. *Practical criticism: A study in literary judgment.* (New York: Harcourt Brace, 1930). c.f. Gage, p. 997.

> into another; least of all of the passage, in the course of many generations, from an animal of a more simple to one of more complex structure."[14]

One would be hard put to recall the specific courses or textbooks or lectures in which one learned the concepts used in this article. How well would one do on an end-of-course exam on biology or intellectual history? One high school graduate who did not go to college but who had read widely could make out the sense of the passage, although she did not know who Lyell was and had not known that the brain of the embryo goes through the stages described. She had some notion of what "transmutation of one species into another" meant, although she probably could not define "species" precisely. She "knew" that the doctrine being discussed had something to do with evolution, but she could not recall the details of the theory. One would expect a college graduate to provide somewhat more ample contexts for the passage and a chap who quit school at the end of the eighth grade a much more fuzzy one, but it is doubtful that the passage would make any sense at all to anyone who had quit school at the end of the fourth grade.

And yet one must not rush into this latter conclusion. The resources of informal education are so great today that it is conceivable that a curious, intelligent adult who had quit school very early in the game could have furnished his mind well enough to construe the passage, and it would be interesting to compare the associative and interpretive resources of such a person with those who had had greater amounts of formal schooling. It is on these sorts of uses and life situations that we badly need research.

Turn now to another passage, one taken from *Time* magazine for June 19, 1972. *Time* is not a highbrow magazine, but apparently its readers are expected to cope with materials of this level of sophistication.

> According to theory, a supernova occurs after a giant star—substantially more massive than the sun—has exhausted its

[14] E. Mayr, "The nature of the darwinian revolution," *Science*, 1972, 176, 983.

> thermonuclear fuel. The star's distended gases begin to collapse toward its center of gravity, crush together and reheat to incredible temperatures of 100 billion degrees, and then explode in a fiery outburst as bright as a billion suns. Left at the center of the supernova is a tiny (about ten miles across) star consisting of tightly packed neutrons, or a smaller "black hole" —a star so dense that its tremendous gravity prevents even light from escaping. The 1967 discovery of pulsars, since identified as neutron stars, seemed to support this explanation of how stars die. Now, observations of Kowal's supernova may help to confirm it.

The lady on whom I tested the first passage could do nothing with this one. She had never studied astronomy and she could not even approximate the concepts that were being employed.

My subject did better on the following excerpt, also taken from the same issue of *Time*. Her wide reading and experiential familiarity with economic transactions left her with a few conceptual gaps, but she could construct a context sufficiently correct to make sense and correct sense out of the passage.

> Most members of *Time's* Board feel more confident than before that the G.N.P. will rise by $100 billion or so, to some $1,146 billion. "We have lots of headroom and a good head of steam," said the University of Minnesota's Walter Heller. Alan Greenspan, chairman of Townsend-Greenspan economic consultants, reckons that corporate profits after taxes will climb 17% this year. Sales volume is running considerably higher than last year in most industries, he explained, and cost-cutting programs started during the recession are now beginning to pay off. In addition, consumers are doing a lot of "up-trading"—moving up from low-priced autos, for example, to higher-priced, and more profitable, models. The Price Commission is concentrating its scrutiny on larger companies, so their profit increases will tend to be somewhat lower than Greenspan's 17% average, and the gains of small firms somewhat higher. Said Joseph Pechman, director of economic studies at the Brookings Institution: "We are witnessing a very, very strong recovery of profits. In general, business will be doing quite well in the next year or year and a half."

Once more it is important to see the response that persons with greater and small amounts of formal schooling would

give us; what blocks to understanding they would encounter; what definite learnings they could trace to this or that course they once took in school.

READING OF LITERARY MATERIALS

I am not suggesting that research has to start from scratch. Some of the relevant variables have been studied, and they could be used to devise instruments for research on adult responses to literature.[15] Such research might keep an eye out for the way in which adult readers construct or fail to construct the appropriate contexts when reading a work of poetry or prose. Especially important, it seems to me, is the distinction, recognized in a number of the researches, between the intellectual, or what might be called the categorical response, and the response in terms of feeling and imagination. The latter, it seems to me, because of its richness and lack of structure is more likely to be overlooked or avoided in systematic research, and yet aesthetically and practically, it may be the most important response, even more important than the cognitive one.

However, before turning to some examples of the imagic associative type of response one might note that literature is itself a source of imagic and valuational associations that could be used in the response to reading matter other than literary. Aside from the mathematical language of science, almost all linguistic products depend on large auras of meaning that are not precisely referential or logically simple. The scrutiny to which carefully composed political documents are subjected is sufficient proof of this. How much of the response to the linguistic products of everyday life are conditioned by or influenced by the study of literature in school? How much of this influence remains, even though the reader no longer can recollect the details

[15] C. W. Harris. "Measurement of comprehension of literature." *Sch. Rev.*, 1948, 56, 280–289, 332–343; J. R. Squire. The responses of adolescents to literature involving selected experiences of personal development. Unpublished doctoral dissertation, Univ. of California, Berkeley, 1956; F. B. Davis. "Fundamental factors in comprehension in reading," Psychometrika, 1944, 9, 186; H. C. Meckel. An exploratory study of responses of adolescents to situations in a novel. Unpublished doctoral dissertation, Univ. of Chicago, 1946.

of the literature he studied? How many of us, for example, respond to materials about Great Britain in terms of the English literature we studied in school? In my own experience, I am afraid that had my first view of the Forum in Rome or the Lake country in England not conformed to my images of these places as formed by my school studies, I would have rejected the reality as spurious, not the images. There will be no opportunity in this paper to go into this aspect of the study of literature in any detail, and there is a defensible ambivalence about using literature as an aid to the social studies. I would agree that to use literature as a source of information about places, events, and characters of history is not the prime justification for such study. Nevertheless, the way in which the poet or novelist perceived his times, the way he personified the values of his time in images is a legitimate resource for the educated response to social problems and the materials dealing with them.

Imagic Contents

To return to the problem of constructing appropriate imagic contexts for the reading of literature. For example, the following stanza from William Butler Yeats' "Sailing to Byzantium"

> An aged man is but a paltry thing,
> A tattered coat upon a stick, unless
> Soul clap its hands and sing, and louder sing
> For every tatter in its mortal dress

depends for its effect and understanding on images with many layers of meaning. "A tattered coat upon a stick," "Soul clap its hands and sing," "mortal dress" are figures of speech, not to speak of the odd hypothesis that if soul clap its hands and sing, an aged man is more than a tattered coat upon a stick. Naming the figures is not enough; paraphrasing them gets closer, but not close enough. It is a matter of the extent of the imagic store; of the type of imagery the person favors; of the fluidity of the images.

In this connection we could do with explorations—at first on individual case studies—of synaesthesia. How much visual imagery, for example, is needed for the Yeats sam-

ple? How much for Gerard Manley Hopkins' ("Pied Beauty")

> Glory be to God for dappled things—
> For skies of couple-colour as a brindled cow;

or Archibald McLeish's ("Ars Poetica")

> A poem should be palpable and mute
> As a globed fruit

Would the *related study of the various arts* increase the imagic store? Would it make a richer imagery available for literary stimuli?

What happens to an image over time as it is stored in experience? John Livingstone Lowes says that 30 years after reading Oliver Wendell Holmes' *Autocrat at the Breakfast Table,* he recalled a passage from it:

> Put an idea in your intelligence and leave it there an hour, a day, a year without ever having occasion to refer to it. When, at last, you return to it, you do not find it as it was when acquired. It has domiciliated itself, so to speak, . . . entered into relation with the whole fabric of your mind.

But when Lowes recalled this passage, it was in the form of something "germinating and expanding . . . with white and spreading tentacles, like the plant which sprouts beneath a stone."[16]

I have in previous years done informal experiments with the imagery evoked by words of Latin origin. Although some college students who had studied Latin did not cite the image of "breathing across" in response to the word "transpire" or "breathe with" in response to "conspire," I have never encountered a student who had not studied Latin who named these images as responses. If English poetry relies on imagery based on Latin (elephant endorsed with towers), then throwing out the study of Latin because it did not transfer to the growth of English vocabulary should have been accompanied by discontinuing Milton's poetry.

In the same vein we could do with some systematic studies on the relation between the images presupposed

[16] J. L. Lowes. *The Road to Xanadu: A Study in the Ways of the Imagination* (New York, Houghton Mifflin, 1927).

by literary works that have their origin in the Bible or Greek and Latin mythology.

Although imagery is a prime factor in the associative uses of schooling, it is not the only one. Suppose we come across the expression "Elizabethan drama." This phrase is a place holder, so to speak, for a host of associations: Queen Elizabeth, Henry VIII, a ruffed collar, Mary Queen of Scots, Shakespeare, knights, battles, the Tower of London, executioners, etc. Some of these associations are imagic, to be sure, but they may include clusters of ideas from history, social theory, and religion. "Henry VIII" would be hard to surpass as a freight train of associations: historic, erotic, religious, and political.

We have had little good to say about storing the mind with stuff—facts, figures, words, definitions, rules—but before thinking can get under way, raw material roughly classified is needed. How many of the responses to adult situations (reading, discussion) betray a sheer poverty of associations? And how often is the poverty of the cognitive kind (ignorance of a relevant fact or generalization) and how often is it moral or aesthetic poverty (inability to retrieve an example, a character, a situation from literature or another art)? One would be much surprised if adults with differing amounts of formal schooling did not reveal differences in their storage and retrieval potentialities.

RESEARCH INTO TACIT KNOWING

Research is needed to throw light on the hypothesis that for the interpretive and associative uses of schooling, retention of the learned material in the form in which it was learned is not essential, an hypothesis that runs counter to the dogma or slogan "Down with nonbehavioral objectives."[17]

The replicative use of schooling does suppose that what was learned will be used *as* learned, hence a simple recall test of materials studied in school gives a fair measure of the probable replicative use. But, if, as has been pointed out, retention of the content of a subject diminishes with time, then the inference will be drawn that this content can be neither applied nor used associatively or interpre-

[17] H. S. Broudy, "Can research," *op cit.*

tively. Because the application of knowledge requires an intimate familiarity with a domain of phenomena, I would hesitate to press to claim that one can *apply* chemical principles or facts which one cannot now recall.

I have discussed elsewhere the role that Michael Polanyi's notion of tacit knowing might play in explaining how we know what we can't tell.[18] It is suggested that a school input interacts with other inputs so that the subject has at his disposal schemata (cognitive, evaluational, and imagic) by which he interprets a task. The input, e.g., the facts and principles of chemistry, or the lines of a poem, or an episode in history, when learned, was the direct target of learning and knowing. Later these learnings became clues to understanding a task; the task is the target and the chemistry facts and principles are clues which are meaningful only as they help make sense of the task as a whole. At this stage one knows the chemistry facts *tacitly, peripherally, not centrally* and *focally*. If we concentrate attention on the facts, they become central again, but their relevance to the task becomes peripheral. Attention switches can make any content tacit or focal, but no content can be both tacit and focal at the same time and in the same way.

Gestalt psychology gives a fair amount of evidence that this conjecture is worth entertaining, but we lack knowledge as to how school inputs function tacitly in nonschool situations; we do not know what concepts, what facts, what rules can best serve as clues to interpretive use. And this is our problem in choosing material for the curriculum of general education. For on this view, general education is the interiorization of that set of clues which in time provide schemata for interpreting the widest possible array of life situations. At the moment I know of no way of getting at these processes other than by comparing adults with varying amounts of formal schooling on their responses to a wide variety of reading materials; to locate the blocks to interpretation, and to hypothesize about the kind of school input that would have removed these blocks. For this purpose real materials at various levels of sophistication should be used, and the subjects should be allowed to

[18] H. S. Broudy, "The life uses," *op. cit.*

verbalize freely either by interview or in discussion. An analysis of tapes or essays may reveal more in the initial stage than objective tests of comprehension or attempts to translate the responses into objectively observable behaviors.

SUMMARY

It should be clear that the style of research proposed here is simply a variation of the Thorndike type of inquiry into life uses of schooling. However, it differs from it in several ways: One difference is that the goal of the research is not to set norms for schooling based on the usage of the "average" man, but rather to find out the difference in use by people with various amounts of formal schooling. Another difference is that the research would emphasize the interpretive and associative uses of schooling in general and language in particular, rather than the replicative and applicative uses. Finally, it would not begin with the dogma that schooling functions in life by being replicated as learned, but would try to tease out what happens to these learnings in the course of time. Is it perhaps the case that so far as interpretive and associative uses are concerned, school learning works best if no longer retained as learned—that perhaps selective forgetting is a necessary condition for creating the interpretive schemata by which we achieve the understanding of our experience?

Some Types of Research on Response to Literature

GUNNAR HANSSON
Gothenburg University

It is now exactly 20 years since I started fumbling around in the field of research which we now call "response to literature." At that time I was virtually alone in Sweden and in Scandinavia working in this field. There was no research to start out from; there were no set problems to develop; and very few of the pertinent international books and research reports had found their way into the Scandinavian university libraries.

So after having worked a couple of years in Sweden, I went to the United States in 1955 to catch up with what had been done in this country. After all, one of the pioneers, I. A. Richards, had been teaching and working in the United States for many years, and his ideas had proved exceedingly fruitful among the New Critics, as well as in people's thinking of literature and appreciation of literature on the whole.[1] Therefore it was something of a disappointment when I discovered that even in the United States very little research related to "response to literature had been done." But I spent an unusually stimulating year here, read yards of interesting books in Harvard University Library, and even made some investigations among school and university students. Unfortunately there was never "world enough and time" to work these things out properly and write them down in a report.

During the last 10–15 years things have started happening, however. In Scandinavia, in the United States, in En-

[1] I. A. Richards. *Practical Criticism*, London Harcourt 1929.

gland, and in several other countries a lively and deeply interesting research activity is suddenly under way. How it all came about and why it seemingly started more or less at the same time in so many places may be difficult to explain. But the activity is there. In Sweden a number of studies have been completed in this field, and several students and teachers are now working with investigations of various kinds. The same can be said about the other Scandinavian countries, although perhaps to a lesser extent than in Sweden.

What I shall try to do here is to illustrate some different kinds of research on the teaching and appreciation of literature which have been carried out or are being carried out mainly in Sweden. Since my main objective is to illustrate different types of aims, methods, and results, the examples have been chosen irrespective of whether the investigations have been published or not, and also irrespective of whether they are on a large or small scale. What is important is to illustrate methods and ways of handling problems, and I hope some of these methods and ways can be fruitfully applied in planning other and perhaps larger investigations in new pedagogical contexts. Several of the investigations I mention have been presented in printed or mimeographed reports. I shall not refer to any names or titles, however, since my main objective is to illustrate the design of the investigations, the problems they try to deal with, and the types of results and conclusions that can be drawn. I shall start with some investigations which deal in general with response to literature, then present some studies of the effects of teaching literature, and finally discuss some efforts to study problems of evaluation in the teaching of literature.

RESPONSE TO LITERATURE

The so-called "protocol" method, where readers are asked to describe their interpretations, experiences and evaluations in written reports, is a well-known and often used method. I. A. Richards used it for his famous books in the 1920's and later, and I used it in my first Swedish investigations in the 1950's. In Richards' *Practical Criticism* all the readers were university students, who studied the poems under the same external conditions. In my first Swedish investigations this simple design was modified in three

respects.[2] Four different categories of readers participated: 1) university students studying literature 2) university students studying psychology 3) college students (age 18) 4) first year pupils at a "people's high school" (adult continuation school, most students age 18–19). These four categories of readers were divided into two or three groups, studying the same poems but under different conditions. Finally attempts were made to influence the individual interpretations and experiences of the poems by group discussions or by letting the readers study different kinds of literary criticism (appreciative or critical reviews, biographical information, historical background, analyses of the meaning and structure of the poem). Only one poem was handed out at a time. The readers were allowed to keep it for a week; then they were asked to comment freely upon it in writing. When the first series of protocols had been collected, either a group discussion took place or the readers were given a written analysis or commentary on the poem to be kept for another week. Then they were asked to write a new protocal and describe if and how their first interpretations and opinions of the poem had been influenced.

This design—which is also very simple, of course—made it possible to present and discuss the students' responses in several dimensions. Thus wide variation in the ability to recreate the poems in rich and personally involved experiences could be demonstrated among the readers, not only between individuals but also, and in some respects especially, between different groups and categories. These differences seemed to be related to factors such as education, general maturity, social background, reading habits, knowledge of literary conventions, critical procedure and principles of evaluation. Some differences were fairly regularly discernible in the protocols from male and female readers.

The attempts to influence the readers' experiences of the poems by means of reviews, analyses and group discussions gave rather limited results. Most of the readers

[2] G. Hansson. *Dikten och läsaren. Studier över diktupplevelsen* [with Summary in English]. (Stockholm 1959); (2nd edition, Stockholm 1970 [No Summary in English]).

could add some details and nuances to their experiences, but more radical changes were few and not always for the better. The results could here be summed up in four main tendencies: the intellectual understanding of a poem could often be enriched and diversified, while the emotional qualities of the experience remained more or less unchanged; shallow and impersonal experiences could be influenced if the reader was not wholly indifferent to the poem; strongly personal experiences were not open to outside influences except for details; well-founded and well-defined critical opinions were hard to change. From an individual point of view, the richest and most personal experiences proved to be the most fragile and susceptible to influences: highly sensitive readers often reacted strongly against factual interpretations, biographical and historical background information, and detailed structural analyses of the texts.

Individual Differences in Response

There is nothing sensational in the demonstration of, for instance, individual and group differences in responses to literature. I. A. Richards shocked the literary world when, almost 50 years ago, he effectively demonstrated how large these differences were among "England's best educated students." We are not shocked; but we tend to forget. People who claim to work "objectively" with literary texts tend to train themselves and others to forget about differences in interpretations and responses. Even teachers are trained to forget about them. And new generations of readers bring new differences to the texts. Therefore, there is still an urgent need for simple descriptive studies, demonstrating to the teachers, and to those who teach the teachers, what happens when groups of young people respond to literature. Our knowledge in this field is still deplorably limited.

The Protocol Method

The protocol method has the immediate advantage of giving interesting and easily available information. When the protocols have been worked through, quoted in illustrative parts and presented in a narrative context, they provide knowledge that is often appreciated by teachers and other persons interested in literature. The knowledge derived from the protocols seems to correspond to their own

way of thinking, talking and writing about literature, and the information they get about the responses and opinions of their students is also of a kind which they find useful for their work in the classroom. Thus the protocols provide information which is both varied and close to reality. It is a useful and important method, especially in the early stages when we are working towards general surveys of the research area, and when we are trying to state hypotheses and define problems which can fruitfully be attacked with more refined methods.

Anyone who makes use of such protocols will soon realize their limitations, however. They raise a number of methodological difficulties. They are often incomparable entities which have to be put into all kinds of Procrustean beds before they can be placed and handled on a common basis. The way to present such protocols in a research monograph will have to be more of an epic description than of a synthetic report of facts. The information that can be derived from these protocols is also very closely dependent on the students' ability to verbalize their experiences and opinions. This ability is in turn closely dependent on their knowledge of a language and a terminology for literary analysis and description which they may or may not have acquired at school or in the society where they have been brought up. What we find in the protocols is thus in an eminent sense of the term an "expressed response": it is by no means the response itself but an acquired way of describing and communicating some aspects of the literary experience that has actually been there.

The expressed response is certainly a legitimate and important target for our research efforts. If the "protocol"-method is combined with the excellent classification system which Alan Purves has presented in his *Elements of Writing about a Literary Work*, or with some similar system based on the principles of content analysis, at least part of the information in the protocols can be handled more systematically.[3] Yet we cannot get round the fact that the protocols will give us more information of what the students have learnt to describe and express in words than of

[3] A. C. Purves and V. Rippere. *Elements of writing about a literary work* (Champaign, Ill. NCTE; 1968).

what they have actually found in the texts when studying them. If we want to get closer to the actual response, we shall have to look for other methods.

The Semantic Differential

One way of getting closer to the response might be to use verbal scales of the kind that Charles Osgood and his colleagues have worked out for their "semantic differential."[4] The method is now well known and has been used in hundreds of investigations on both linguistic and other kinds of material. To illustrate its application to problems in response to literature, I shall briefly outline the design of a study where the standardized scales of the "semantic differential" were not used but where basic ideas were derived from Osgood's work.[5]

The study deals with a Swedish poem of the beginning of the 19th century, a subtle and not easily interpreted poem of the romantic period. There has been extensive discussion among scholars and critics about the meaning and proper interpretation of this poem. In this debate there are many statements that clearly refer to historical circumstances: the history of ideas, motifs and style, the author's life, and the like. Statements of that kind are not at all dealt with in the study. On the other hand there are many statements that explicitly or implicitly refer to the response of somebody who *now* reads the poem, statements like the following: "Stanza two is the most tragic," "The compassion is most moving in the last two lines," "The suggested tension in the first lines is held in suspense and then reaches its climax in the last three lines."

When we are confronted with statements like these we can—and we must, if we want to understand them fully—ask questions: *Who* feels like that when reading the text? *Who* is responding like that? And when we have started asking such questions, we can easily go on: Do students in elementary school respond like that? Do college students respond like that? Do literary critics and scholars unanimously respond like that? These questions,

[4] C. E. Osgood, G. J. Suci and P. H. Tannenbaum. *The Measurement of Meaning* (Urbana, Ill. University of Illinois Press 1957).

[5] G. Hansson. *Dikt i profil.* (Göteborg Akademiforlaget-Gumperts 1964).

which are not mocking but represent serious interpretations of incomplete, elliptic statements about the poem, can then be stated as a series of hypothetical statements about the responses of specific groups of readers. The original statements were certainly not intended as such hypothetical statements, but for research purposes (and perhaps also for purposes of intellectual clarity) we have interpreted them, completed the ellipses, and reformulated them as hypotheses.

When we have a number of such hypotheses, we can construct a series of verbal scales intended to register the meaning or response qualities which the hypothetical statements are referring to. If we present the scales to specific groups of readers and ask them to mark their judgments on the scales, our results will show whether the readers responded to the text in a way which confirms the hypotheses or not. For most literary texts it would probably take a large number of scales to represent all qualities for which we could find or develop hypotheses; but we can choose among them, picking out for instance those that are crucial in the critical debate, or most important in a pedagogical context.

Response Scales

This is what was done in the Swedish study which I am referring to. In all, 25 bipolar 7-point scales were constructed, of the following type:

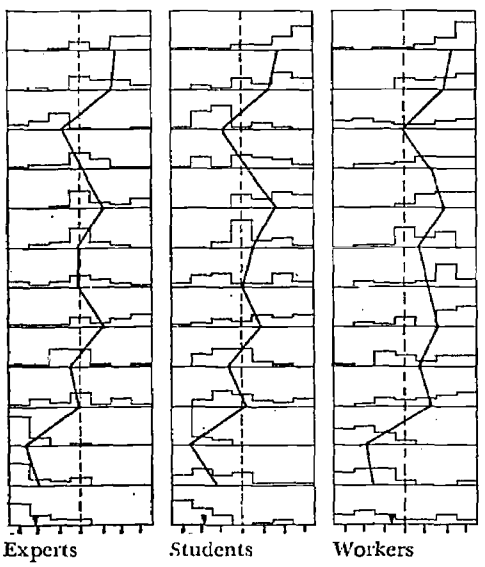
TRAGIC 1 2 3 4 5 6 7 HAPPY

Other scales included Hopeful-Hopeless, Cold-Warm, Restrained-Passionate, Simple-Complex, Quiet-Loud, Fast-Slow. The poem was divided into 12 parts, each consisting of 2 lines which, apart from their contextual relations, were complete meaning-carrying units. Three group of readers took part: experts (scholars or teachers of literature), university students studying literature, and skilled workers with only 7 years of compulsory education in their childhood. All persons participating had to judge each one of the 12 parts of the poem on the whole set of 25 scales, and finally also the whole poem. Figure 1 shows means and distributions for the three groups of readers on three of the scales.

Figure 1

Means and distributions for three groups of readers on three scales:
1. TRAGIC–HAPPY 2. RESTING–MOBILE 10. SIMPLE–COMPLEX

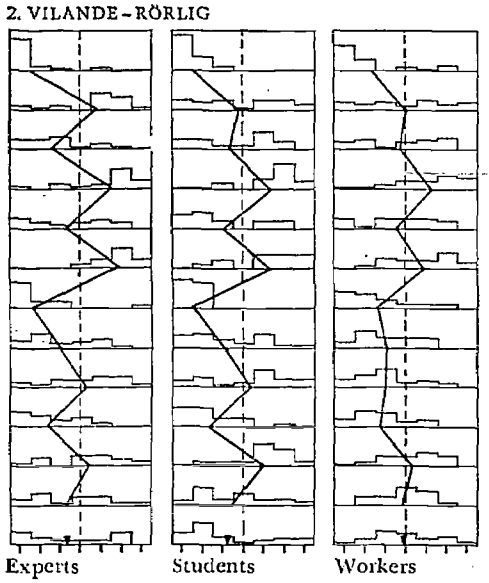
1. TRAGISK – LYCKLIG



1a Skön, med lågande hy och slutna ögon,
slumrar herden så ljuvt i månans strålar.
1b Nattens ångande vindar
fläkta hans lockiga hår.
2a Stum, med smäktande blick och våta kinder
honom Delia ser från eterns höjder:
2b nu ur strålande charen
svävar hon darrande ned.
3a Och av klarare ljus, vid hennes ankomst,
stråla dalar och berg och myrtenskogar.
3b Utan förerska spannet
travar i silvrade moln.
4a Herden sover i ro: elysiskt glimma
i hans krusiga hår gudinnans tårar.
4b På hans blomstrande läppar
brinner dess himmelska kyss.
5a Tystna, suckande vind i trädens kronor!
Rosenkransade brud på saffransbädden
5b unna herden att ostörd
drömma sin himmelska dröm.
6a När han vaknar *en gång*, vad ryslig tomhet
skall hans lågande själ ej kring sig finna!
6b Blott i drömmar Olympen
stiger till dödliga ned.

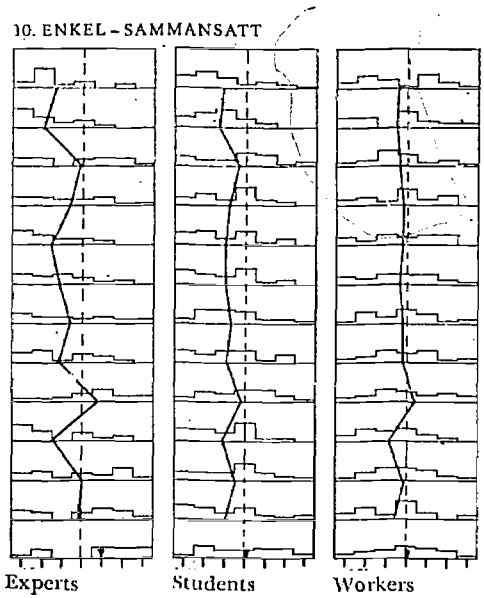
Whole Poem

2. VILANDE – RÖRLIG



1a Skön, med lågande hy och slutna ögon,
slumrar herden så ljuvt i månans strålar.
1b Nattens ångande vindar
fläkta hans lockiga hår.
2a Stum, med smäktande blick och våta kinder
honom Delia ser från eterns höjder:
2b nu ur strålande charen
svävar hon darrande ned.
3a Och av klarare ljus, vid hennes ankomst,
stråla dalar och berg och myrtenskogar.
3b Utan förerska spannet
travar i silvrade moln.
4a Herden sover i ro: elysiskt glimma
i hans krusiga hår gudinnans tårar.
4b På hans blomstrande läppar
brinner dess himmelska kyss.
5a Tystna, suckande vind i trädens kronor!
Rosenkransade brud på saffransbädden
5b unna herden att ostörd
drömma sin himmelska dröm.
6a När han vaknar en gång, vad ryslig tomhet
skall hans lågande själ ej kring sig finna!
6b Blott i drömmar Olympen
stiger till dödliga ned.

Whole Poem



1a Skön, med lågande hy och slutna ögon,
slumrar herden så ljuvt i månans strålar.
1b Nattens ångande vindar
fläkta hans lockiga hår.
2a Stum, med småktande blick och våta kinder
honom Delia ser från eterns höjder:
2b nu ur strålande charen
svävar hon darrande ned.
3a Och av klarare ljus, vid hennes ankomst,
stråla dalar och berg och myrtenskogar.
3b Utan förerska spannet
travar i silvrade moln.
4a Herden sover i ro: elysiskt glimma
i hans krusiga hår gudinnans tårar.
4b På hans blomstrande läppar
brinner dess himmelska kyss.
5a Tystna, suckande vind i trädens kronor!
Rosenkransade brud på saffransbädden
5b ur.na herden att ostörd
drömma sin himmelska dröm.
6a När han vaknar en gång, vad ryslig tomhet
skall hans lågande själ ej kring sig finna!
6b Blott i drömmar Olympen
stiger till dödliga ned.

Whole Poem

On the whole there is a striking similarity in the diagrams for the three groups of readers. Since the poem is difficult to interpret and the qualities judged by the readers are subtle and not easily accessible, there is no doubt that marked differences would have been found in written protocols, if this method had been used. One conclusion that can be drawn is that the *passive* ability of the less educated readers to notice and judge linguistic, literary and experiential qualities is much more developed than their *active* ability to verbalize their interpretations and experiences in a written statement. We might also say that the readers were given in the scales an instrument by which they could make distinctions in their responses and in their use of language when reading the poem. With this instrument the less educated readers could make more or less the same judgments as the more educated readers could. There is one exception to this, however. On scales that register *formal* qualities, like Simple-Complex (see diagram no. 4 in Figure 1), Fast-Slow, Vibrant-Static, the less educated readers have not been able to make as diversified judgments as the more educated. The differences in this respect can easily be explained. The judgment of such formal qualities presupposes not only ability to give mean-

ings to words and sentences but also practice in literary analysis and knowledge of critical terminology. These things are important in school and university courses, and form a kind of expert knowledge, but in society at large they are of secondary importance.

The confirming or refuting of the hypotheses derived from statements made by scholars and critics was not a main objective in this study. But it is evident that the judgments on the scales could be used to demonstrate whether there had been present in the responses of the readers such qualities as the scholars and critics had attributed to the text. In the next stage it could be demonstrated whether there had been such relations between qualities as certain critics and scholars had suggested. Thus the data could be analyzed to demonstrate structural relations in the responses and interpretations of the readers.

Another conclusion that can be drawn from this and other similar studies is that for literary and probably most other aesthetic investigations, unipolar scales are more useful than bipolar ones. The main reason for this is that bipolar scales presuppose *either-or*: either *tragic* or *happy*, either *cold* or *warm*. But many literary works and other aesthetic objects demand *both - and* in the response: both *tragic* and *happy*, both *cold* and *warm* at the same time, or at least in the same response. Instead of excluding each other, different qualities are often put together in sharp contrasts. Therefore, to avoid forcing a reader to choose between two opposites on a bipolar scale, we should often include two unipolar scales, one for *cold* and another for *warm*, for instance.

Scales as Teaching Devices

From studies of this kind there are other paths leading into the practical teaching of literature. The task of putting marks on the scales has often proved highly stimulating and motivating to the students, especially perhaps to students who are not otherwise particularly interested in the study of literature. When they have made their marks on the scales, they have at their disposal a linguistic material which they have produced and observed and structured on their own. They are often eager to discuss this material, both to make their own markings clear and to know why their friends have made other markings. This linguistic

material is not handed over to them by the teacher, or by the textbook, or by their "better" friends: each one of them has produced his own material. This is a very fruitful teaching situation, which can be made use of in many different ways. Generally, it can be used as a starting point for an active, student-centered training of basic linguistic abilities, which are a prerequisite for the appreciation of literature. In a classroom discussion, for instance, where the students are of different opinions as to the meaning of some words or the interpretation of a stanza, two or three "hypotheses" can be formulated on the spur of the moment. A few scales can be constructed, and a small investigation can be improvised in the classroom. The results will show how all members of the class have judged the meaning qualities that were at issue, and they have all contributed to the clarification of a problem of interpretation.

A teacher recently described in an article one way—and in his own opinion a successful way—of using verbal scales in practical teaching.[6] His pedagogical aim for a two-hour lesson was to demonstrate to his students how the individual and his personal problems had been the central theme in Swedish poetry after the Second World War and up to about 1965, and how this private, psychological orientation had been replaced by a more collective, society-oriented, political consciousness. All teachers know how little effect they may expect in the classroom if they stand up and tell the students about such a change in the literary climate, and then hopefully ask them to make use of this information when reading a number of texts from the period. This teacher constructed seven bipolar scales with 5 points on each. One of the poles represented the early 60's, the other the late 60's, e.g. *Complicated–Uncomplicated, Extroverted–Introverted, Abstract–Concrete, Authentic–Fictitious,* The students were asked to apply these scales to a number of poems, some from the early 60's others from the late 60's. When the markings had been collected in simple statistical tables and the poems had been dated by the teacher, the students themselves had found out and dem-

[6] R. Yrlid. "Diktanalys med tvapolsskalor. Ett tillämpningsförsök," *Svenskläraren 1972:2,* p. 18–21.

onstrated significant differences between the two groups of poems.

This design was used in practical teaching, but it could easily be turned into a research design by formulating the pedagogical aim as a research problem or as a series of hypotheses to be tested empirically. It is important, not to say necessary, that investigations of this kind are governed by hypotheses or distinct problems. If we just start using the "semantic differential" or some other set of scales in the field of response to literature, hoping that something will come out of it, we shall probably not make much progress. The "semantic differential" is explicitly intended to measure general semantic dimensions, and it seems to be an excellent instrument for this purpose. But such general dimensions do not help us much in the teaching of appreciation of literature. What we need are hypotheses and distinct problems which are derived from crucial and recurring teaching situations, and then we need scales which measure in the dimensions outlined by such hypotheses and problems. There are many ways of getting down to crucial problems: one way is to ask teachers about their experiences and expectations in teaching. I shall illustrate this by outlining the design of a study that has been carried out in Sweden.[7]

Scaled and Written Responses

For this study were chosen four texts which are very often studied in the Swedish schools. In the first stage a questionnaire was sent out to a sample of teachers, asking them to describe in detail their teaching practice when studying these particular texts in class: what kind of problems (history of literature, author's life genre, style, etc.) they tried to deal with in the study of the texts, how they generally introduced and carried through as analysis of the texts, what they expected from the students, what in their experience was easy or difficult for the students to master in the analysis and study of the four texts. The statements made by the teachers were used for the formulation of

[7] B. Hansson, "Studenter, gymnasister, realskoleelever och grundskoleelever tolkar och värderar Edith Södergrans dikt 'Landet som icke är'," Mimeo, Göteborg 1971. B. Hansson, "En uppläggning av undervisning om litterär värdering för gymnasieskolan," Mimeo, Göteborg 1971.

hypotheses and problems, and thus for the choice of dimensions in which to construct the scales.

In all, 14 scales were used for each text in this study. Four of the scales were common to all four texts: they were chosen so as to register qualities which, according to statements by the teachers, were relevant and important to the study of all four texts. Six of the scales were chosen so as to be relevant to one of the texts; thus there were four different sets of these six scales. The remaining four scales were intended to register qualities which the teachers had *not* mentioned as relevant in the study of the particular text. Besides applying the set of 14 scales, the students had to write "protocols" in which they formulated their opinion on one or two problems of interpretation or characterization. These problems had also been chosen on the basis of statements made by the teachers. Finally the students were asked to evaluate the texts and to give as many reasons as they could for their evaluations. Three groups of students, representing three stages in the education system, took part in the study: grade 9 in the compulsory school (age 16), grade 2 in the "gymnasium" (age 18), and university students studying literature (age 20–25). All three groups were given exactly the same assignments.

The main result was the same as has already been illustrated in Figure 1: a striking and certainly not expected similarity between the three groups in the means and distributions of the markings on the scales. But in this case there were conspicuous differences in the written protocols from the three groups: in the amount (number of words and sentences) of what they had written, in their ability to use descriptive and expressive language to communicate what they had found in the texts, and in the content of their interpretations. The differences in their interpretations can be illustrated by Table 1, where inter-

Table 1

	Interpretation			
	A	B	C	Mis.
Compulsory	32%	22%	22%	24%
Gymnasium	70%	16%	5%	9%
University	70%	9%	1%	20%

pretation A represents the "expected" interpretation, and B and C are two variant interpretations often proposed by the students.

We can see in the table that the students in the compulsory school arrived at three different interpretations (apart from a group of miscellaneous interpretations), and that they divided themselves fairly evenly among these three. Two of these interpretations were then gradually abandoned by the students in the gymnasium and at the university. When these differences in the verbalized interpretations are contrasted with the similarities in the markings on the verbal scales, we can of course suggest several explanations. One might be that the scales were too simple or too crude for the students: they did not discriminate properly. I do not think that is a plausible explanation; but to be certain we need more research. Another explanation, which to me seems more plausible, would be that the 16-year-old students in the compulsory school have the basic linguistic ability to give meanings to the words of the text and to make distinctions between meaning qualities structured by the text. What they lack to a certain extent is the ability to organize these meanings and qualities into a coherent whole which they can put into words and communicate as an interpretation or an analysis of the text. This latter ability needs training; and that is what is being done in the higher stages of the educational system.

EFFECTS OF TEACHING LITERATURE

Studies of what happens during a period of literature teaching are few, at least in Scandinavia. The literature teacher hopes to achieve a number of positive effects, but we know very little of what he really achieves, either in his students' appreciation of particular texts or in their attitudes to literature on the whole. The few studies that have been made have all demonstrated positive effects, but at the same time all of them have pointed out some clearly negative effects which every literature teacher would certainly like to avoid, if possible. It is evident, for instance, that many students strongly oppose being given solid, detailed historical or biographical background information to poems or other works for which they have a personal liking. Many students also react, sometimes passionately, against advanced technical analyses of a kind which especially

the new critics have cultivated. In a country like the United States it would be a worthwhile subject for research to demonstrate the pedagogical effects of these kinds of analysis, which evidently have been and still are much favoured in higher education.

Conflicting Aims in Teaching

The literature teacher has various aims in his teaching: to give his students knowledge of individual works and authors and of the history of literature, to develop their ability to understand and appreciate literature on their own, to implant positive attitudes to reading on the whole, to encourage them to go on reading after they have left school. These and other aims may conflict with each other in a teaching situation, and then the teacher may have to sacrifice something to achieve something else. Every teacher knows the situation from his own experiences; but what happens in the class, and why does it happen?

What kind of answers we can expect to such questions may be illustrated by a small study where one group of students studied, analyzed and discussed a short story in class, while another group were just given a few written statements about the author and then studied the story individually.[8] The students' attitudes to the story and its author, their understanding of the story on a fairly elementary linguistic level, and their interpretations on a more literary level were registered. What happened was that the students who analyzed and discussed ths story in class acquired a more uniform understanding of the text and reached more uniform interpretations. So far the teaching could be said to have led them in the "expected" direction. But at the same time the attitude of these students had grown more negative, and they felt less keen on reading more texts by the same author than did the students who not had been taught in class. These two pedagogical effects were produced in the same teaching situation. A more unified interpretation of a text can be a desired effect: not in itself, I would say, but if the students' ability to interpret texts has been forwarded in a way which the school, the teacher, or the students themselves later on will find useful. The negative attitudes are certainly not de-

[8] G. Hansson. "Med undervisning och utan," *Svensklärarföreningens Årsskrift 1970,* p. 45–60.

sired, and the question is whether we can defend having them included in the bargain.

Effects of Teaching Methods

In later studies the effects of different ways of teaching literature have been compared. I shall briefly outline the design of one of them.[9] This study focused on two texts from the 1890's, an interesting period in the history of Swedish literature, when realism and naturalism were replaced by a romantic and aesthetic movement. Three groups of students took part. One group just read the two texts individually but did not analyze or discuss them at all in class. In the second group the teacher introduced the two texts by informing the students of the historical background and of the intentions and aesthetic aims of the authors. The analyses and discussions of the texts were then carried out in the framework of this background knowledge. In the third group there was no introduction of this kind but a free discussion, where the students could ask questions and raise problems which they wanted to discuss or be informed about after having read the texts. The teacher acted as a chairman in the discussions, but he did not raise any problems or provide any information unless the students asked for it. There were four classes in each of the three groups.

In the first stage the students just read the two texts a few times, and immediately afterwards they answered a number of attitude questions: their attitudes to the texts and to having them further analyzed and discussed in class. Then followed the teaching in the two experimental groups. Since teaching literature undoubtedly to some—and perhaps to a large—extent is dependent on the teacher's personality and other individual qualities, four teachers had agreed to play two different roles in two otherwise comparable classes: in one class they taught "historically," in the other they conducted a free discussion. It may be difficult for a teacher to play two different roles without any bias on one side or the other, but to a large extent the effect of the teachers' personal qualities was probably controlled that way. After the teaching the students had to

[9] A. Fredholm, "Gymnasisters litterära värderingskriterier, "*Svensklärarföreningens Årsskrift 1970,* p. 61–100.

answer a number of questions intended to register 1) their knowledge of some facts about the period and the authors; 2) their understanding of the texts; 3) their response to more subtle and literary aspects of the texts; 4) their attitudes to reading more texts by the same authors; and 5) changes in their response and attitudes because of the teaching.

Comparisons between the two experimental groups and the control group showed that both kinds of teaching were effective and on the whole gave positive results. Comparisons between the two experimental groups showed that the "historical" teaching in this case had been more effective. The students in these classes had acquired more knowledge of the period and of the authors, and they showed a "better" understanding of the texts, both on an elemetary level and on a more subtle and literary level. On the whole these students had also somewhat more positive attitudes to reading other texts by the same authors.

What is "better"?

The word "better" is of course problematic in this context: what is a "better" understanding, a "better" response or a "better" interpretation? But these problems can be handled in objective research. If we do not think of "better" as *right* or *wrong*, but as *expected* interpretations and responses, we can always define what the expected response is, and why we expect it. Our expectations will vary with different teaching goals and different teaching situations, but if we know what we are aiming at, we can always define and defend them. If the teaching is built on historical information, as was done here in one of the experimental groups, we can for instance expect a particular response because the historical information was intended to make the students notice specific qualities in the text. If the teaching aims at getting the students involved in a particular text, we can expect (and accept as "good") widely different interpretations if only they are personally satisfying to those who have proposed them. When the expected responses have been defined in this way, we can for research purposes easily make a numerical index, showing how far the teaching we are studying has brought an individual student or a group of students.

PROBLEMS OF EVALUATION IN THE TEACHING OF LITERATURE

Questions of evaluation are central in all response to literature. It is hard to imagine a person reading a work of literature without being full of evaluative reactions while he is reading and without ending in an evaluative summing up of what he has read. His evaluation may be more or less thought out, more or less supported by reasons, but it will be there. In the teaching, questions of evaluation have been badly neglected on all levels, however: in schools and universities, and in the education of teachers. In the Swedish curriculum plans, which are unified and operable for all schools in the country, problems of evaluation have recently been introduced as a point to pay special attention to in the teaching of literature. But that is only very recently. Supposedly these problems have been just as much neglected in the United States and most other countries as they have been in Sweden. I have been told that so-called "objective criticism" has held such strong sway in the United States that questions of evaluation have been practically banned from the teaching of literature in most schools and universities, the reason being that evaluations are regarded as purely subjective and therefore have no place in objective teaching. And yet teachers and students have certainly continued to evaluate and to argue, at least silently, for their positive or negative opinions of their reading.

Taste and Reasoned Taste

As has often been pointed out, taste is at least partly withdrawn from rational arguments. Taste can be trained, it can be developed and refined by abundant reading and certainly also by teaching, but taste itself cannot be taught. The immediate feeling for what is valuable is probably based in attitudes and dispositions which are central parts of the reader's personality and which can be changed only through deep crises or long passages of time. But in a different sense *de gustibus non est disputandem* is all wrong: just because taste is so different, it is more important here than in most other areas that we train ourselves to state our own reasons and to understand other people's reasons for their evaluative opinions. When we state our reasons, we are using rational principles of evaluation and critical frames of reference. Such principles and frames of reference often operate within larger systems of ideas—philo-

sophical, moral, political, religious—and within these systems different principles of evaluation can both be formulated and put to rational tests. It is also possible—in teaching, for instance—to make an individual reader conscious of how his principles of evaluation function, and of other principles which he may not be familiar with. It is also possible to show him when and where some principles are not logically coherent, when they are or are not applicable, and when they lead to unacceptable consequences.

This area, the use of principles of evaluation and critical frames of reference by teachers, students and other readers, seems to me to be a very important area for our research and teaching efforts. One of the main goals for the teaching of literature should be the development of a critical frame of reference in the students, thus enabling them to make rational judgments of what they are reading, to state reasons for their own choices and evaluations, and to understand the reasons for other people's choices and evaluations.

To do research and teaching with respect to literary evaluations, we must have a fairly complete survey of the criteria which are available in the society in which we live and work. In Sweden there have been made or are being made several studies aiming at such surveys of available criteria. Some of these studies are focusing on pedagogical situations, others are focusing on the use of criteria in literary criticism, both present-day criticism and earlier criticism. I shall illustrate the kind of problems dealt with by some studies focusing on pedagogical situations.

Varieties Of Evaluation

One study used seven literary texts of varying type and quality.[10] These texts were handed out to two different groups of students, who were asked to grade them on a 9-point scale ranging from −4 to +4, where −4 represented one of the worst texts they had read and +4 one of the best texts they had read. When the grading was completed, the students in one group were asked to support their gradings by choosing 5–10 criteria which they agreed with from a list of 17 preformulated criteria. These 17 criteria cor-

[10] L. Skaaret, *"God" och "dalig" litteratur. Gymnasister och deras lärare on tva berättelser* [with Summary in English]. Stockholm, 1971.

respond to Alan Purves's "elements of evaluation" in his *Elements of Writing about a Literary Work,* but they had been simplified to match the students' ability to understand the critical vocabulary. The students in the other group were asked to write down the reasons they had to support their gradings. These reasons were then classified according to the same system of 17 criteria. Finally the students were asked to grade the seven texts once more on a 9-point scale.

Table 2 shows how many of the students in the two groups used each of the 17 criteria one or more times:

Table 2

Criterion no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
free comments	12	25	19	27	11	3	3	22	4	13	27	16	27	8	22	27	15
preformulated	18	32	21	27	27	30	11	31	30	30	31	28	30	31	25	28	20

Several of the criteria were hardly used at all by the students who in free comments stated the reasons to support their gradings. Criterion no. 9, for instance, which is "the author's intention," was used by only four students writing free comments, while it was used by 30 students who were given the preformulated criteria. Table 3 shows how many different criteria each student in the groups mentioned:

Table 3

Number of diff. criteria	0	5	6	7	8	9	10	11	12	13	14	15	16	17	M
free comments	1	3	4	4	5	4	4	3	4	1					8.4
preformulated									4	4	15	5	3	1	14.1

Three students writing free comments mentioned 5 different criteria, four students mentioned 6 different criteria, etc. On the average, students writing free comments mentioned 8.4 different criteria when evaluating the seven texts while students using the preformulated list of criteria mentioned 14.1.

Conclusions from a first small study of this kind will have to be tentative; but they may point out fruitful paths to follow, in research and in teaching. Some tentative conclusions might be that if students are asked to state in free comments the reasons they have for value judgments which

they have made, they can use only a small number of different criteria. If they are made familiar with the whole range of available criteria, they will support their judgments by other kinds of reasons than those they generally make use of. Their range of arguments is almost twice as wide, which means that their judgments are much more diversified. Many students evidently have a passive ability and a willingness to use other criteria than the few they are already familiar with.

Changes of Judgment

In this study a number of students who consistently graded the texts on or near the extremes on the scale were compared with other students who consistently graded the texts near the middle of the scale. Students who graded near the middle of the scale, and then stated their reasons in free comments, did not change their opinion at the second grading of the texts: their vague and undecided judgments had not been changed by continued study of the texts. The students who marked near one of the extremes and then wrote free comments did change their judgments, however; they moved further out towards the extremes at the second grading. Continued study of the texts while they searched for arguments made their already definite judgments still more definite. The students working with the preformulated list of criteria showed just the opposite tendency. The students who first graded the texts near the middle of the scale, placed their markings further out at the second grading. The list helped them to see new aspects of the texts and thus make more definite judgments. The students who first graded near the extremes moderated their judgments by moving their markings towards the middle of the scale. Also these students were helped by the list to see new aspects of the texts, and evidently they realized that their first judgments were somewhat one-sided or oversure, a situation which they now modified.

Students and their Teachers

In another stage of this experiment the students were asked to rank the preformulated criteria, first as to how important they were in their own judgments of literature, then as to how important the students thought they were in the judgments made by experts, critics and teachers of

literature. As can be seen in Figure 2, the differences were considerable. The main tendency is clear. When the students rank for themselves, the most important criteria are the emotional impact of the work, the author's imagination, the moral significance of the work, and the author's sincerity. All these criteria, which pay attention to human qualities in literature, to the significance and import of the message communicated by the works, are placed low down in the supposed ranking by experts and teachers. Instead, a number of purely *formal* criteria, such as form and style, aesthetic order, and symbols and metaphors, are considered to be very important in the judgments of experts and teachers.

Results of this kind may make us wonder about the way literature is taught at school, and what gives the students their opinions of what is more or less important to their teachers. The tendency I have illustrated is confirmed by

Figure 2

Rank order	Ranking by students.	Ranking by experts (supposed by students)
1	* affective evaluation	* form and style
2	* author's imagination	* aesthetic order
3	* moral significance	* author's intention
4	* author's sincerity	* symbols and metaphors
5	* author's intention	* thematic importance
6	* thematic importance	* generic evaluation
7	* originality	* author's sincerity
8	* multifariousness	* multifariousness
9	* symbols and metaphors	* author's imagination
10	* mimetic plausibility	* traditional evaluation
11	* generic evaluation	* affective evaluation
12	* aesthetic order	* originality
13	* form and style	* moral significance
14	* moral acceptability	* moral acceptability
15	* symbolic appropriateness	* mimetic plausibility
16	* traditional evaluation	* symbolic appropriateness
17	* citation of criteria	* citation of criteria

several independent studies: formal criteria such as form and style, aesthetic order, and symbols and metaphors are gradually gaining ground through the educational system, while criteria like emotional impact and mimetic plausibility are gradually losing ground. Do we have to start a debate on "the dehumanization of art" once again?

In the studies I have just referred to the results showed *rankings* of criteria of evaluation according to how important they were considered by groups of students. But the same tendency of "dehumanization" was apparent in a longitudinal study, where results showed the *actual use* of different criteria. In this study, which I have previously mentioned (see page 271 above), the students were asked to write down the reasons for their value judgments. These reasons were then classified according to the same classification system that has been used in the other studies I have mentioned. Students from three different stages took part in the study: last year in the compulsory school (age 16), second year in the gymnasium (age 18), and university students studying literature (age 20–25). On all three stages three criteria made up 60–70 per cent of all criteria mentioned by the students, but the three criteria were not the same. Figure 3 shows the three criteria most often mentioned by the students on the three stages:

Figure 3

Compulsory	1. emotional impact 2. moral significance 3. thematic importance
Gymnasium	1. form and style 2. emotional impact 3. moral significance
University	1. form and style 2. emotional impact 3. symbols and metaphors

The formal qualities are gradually taking over as the educational system progresses towards producing experts, critics, teachers. Does this tendency coincide with the general goals set up for the schools? Is this stressing of the formal aspects of literature really supported by the attitudes and opinions of the students and of their teachers? Would

not systematic training of the students in the use of all available criteria enable them to make more diversified and better supported judgments, and thus make them more successful readers?

The last question especially supports the case for experimental teaching, and for further research. Some recent studies have shown that the number of criteria actually used by students—and by their teachers—is often surprisingly small. Many students and teachers do not use more than three, four or five different criteria in their everyday talking and writing about literature. These criteria can be formulated in dozens of ways, but in an analysis they are reduced to just a few types.

An Experimental Study

A small preliminary study can illustrate how these problems may be attacked in teaching and in research.[11] Two comparable texts were chosen: text 1 was presented to one group of students and text 2 to another group of students. The texts were presented without title or name of the author. The two groups of students were equal in respect of achievement and interest in literature. These students were on a practical line of education, where much time cannot be devoted to literary studies and where the teacher cannot expect much interest in literature. The students were asked to make a value judgment of the text they had read and to state the reasons for their judgment. Then for about a month a few minutes every now and then were devoted to informing the students of different principles of evaluation and to training them in the use of various criteria. When this kind of teaching was finished, the texts were changed over so that the students who had read text 1 now were given text 2, and vice versa. Judgments and reasons for the evaluation of the new text were stated as before.

On the first occasion the students used only 5 different criteria. On the second occasion, i.e. after the teaching, they could use 10 different criteria. In a short period of fairly improvised teaching these students had acquired and learned to use 5 new criteria: the author's intention,

[11] A. Fredholm, "En undersökning av gymnasisters litterära värderingsnormer," Mimeo, Göteborg 1971.

the author's imagination, aesthetic order, citation of criteria, and orginality. On the second occasion the evaluations were more diversified in another way too: 9 students as against only one on the first occasion supported their judgments by referring to both positive and negative qualities in the text. Also the average number of different criteria used by each student increased considerably, from 1.4 to 3.2 in one group and from 0.9 to 2.5 in the other.

These figures are small, and perhaps insignificant in many respects. But they point to development of responses to literature in a direction which it should be a pleasure to define as "expected."

An Experience Preserved: The Video Tapes of the Minnesota-NCTE Seminar

MICHAEL F. GRAVES
University of Minnesota

If one has read the invited papers collected in this volume, one has some idea of the input to the Minnesota-NCTE Seminar on Research. Assuming further that one has read Peter Rosenbaum's introduction to the collection, one realizes that the papers constitute only part of the output of the Seminar. To say this is in no way to minimize the importance of the papers. Without them, there could have been no Seminar. What took place was a conference, not a rap session; and the conference could not have taken place without the scholarship, insight, and direction provided by the papers.

But the sum total of the Seminar was in many ways much more than the papers themselves. To begin with, the papers were not read at the conference. Participants received and read the papers in advance. At the Seminar, the ideas presented in each paper were discussed by a panel and the group as a whole in plenary sessions, further considered by small working parties, and then praised, damned, rehashed, and integrated with other views during the nearly continuous interchange that took place throughout the three days of the conference. The sum total of the Seminar was, as Rosenbaum points out, *an experience*. And its potential for influencing research within the profession lies very much in one's being able to participate in that experience.

To be sure, the week of November 20, 1972, has passed.

One can no longer make reservations for that week in what Rosenbaum seems to remember as that small, isolated motel in the wilderness of St. Paul. But at least a part of the experience still exists. Each of the seven plenary sessions, a final discussion on the Seminar between James Squire and Dwight Burton, and an overview of the entire conference have been recorded on video tape.

Watching the tapes is not, of course, the same thing as being at the Seminar. At the same time, the immediate and direct rendering of reality is extremely forceful. The "blips on the screen" are alive. Harry Broudy's eloquence, which I am told regularly packs Illinois classrooms, is not merely noted; it is present. So too is Ken Johnson's wit, and the drama of Gunnar Hansson and James Britton debating basic assumptions regarding literature and literary response.

But there is more than this to the tapes. Watching the tapes is in some way actually superior to having attended the Seminar. Attention is focused. The "blips on the screen," the fleeting but perhaps important comments are preserved. Something not quite grasped can be replayed and analyzed. The Talbert-Calfee exchange that Rosenbaum refers to as a "breakdown in communication" can be seen in a fuller and perhaps quite different light. The "need for interdisciplinary understanding" can be felt, but so too can the difficulty and challenge of achieving such understanding.

There is a good deal more that could be said about the tapes. But perhaps the general picture is clear. Let me turn now to the task of giving a brief preview of each tape.

The Minnesota-NCTE Seminar on Research in English Education. 35 minutes.

This composite tape includes highlights from throughout the conference and is meant to serve as an overview of the Seminar and an introduction to the tapes. This introduction explains the purpose of the Seminar, outlines the program, identifies the consultants and their papers, and highlights the following major topics: the value of empirical research, the need for interdisciplinary cooperation, new instructional modes, graduate student training, and the relevancy of research.

Plenary Session One:
"Research in English Education: The Troubled Dream."
By DWIGHT BURTON. 55 minutes.

Panel members are Mr. Burton, Walter Petty (participating moderator), Janet Emig, and Herbert Simons. The discussion focuses on what is right, what is wrong, and what ought to be changed in the research in English Education. Topics include the province of research in English Education, the need for better coordination of research, the possibilities of establishing research centers focusing on particular areas, the quality of research in the field, and the quality of the research training graduate students receive.

Plenary Session Two:
"Anthropological Research Models."
By CAROL TALBERT. 60 minutes.

Panel members are Ms. Talbert, Roy O'Donnell (participating moderator), Robert Calfee, and Kenneth Johnson. The discussion focuses on the sorts of research and spirit of research undertaken by anthropologists, the tension between naturalistic and empirical research, and the necessity of educational researchers' knowing the "cultures" within which they work. Topics include the meaning of "ethnicity," the validity of generalizations based on different sorts of evidence, the harm wrought by researchers who have not known the "cultures" of the students with whom they work, and the need for interdisciplinary cooperation.

Plenary Session Three:
"Measurement and Evaluation."
By JEREMY FINN. 50 minutes.

Panel members are Mr. Finn, Doris Gunderson (participating moderator), William Fagan, and James Hoetker. The discussion centers around the sorts of research English educators ought to be doing, the research methodology appropriate to educational problems, and—the training in research provided for graduate students in education. Topics include the necessity of researchers' having logical conceptual bases for their work, the desirability of using multivariate designs and analyses for educational data,

and the need for graduate students in education to receive thorough training in statistics and design.

Plenary Session Four:
"Some Types of Research on Response to Literature."
By GUNNAR HANSSON. 50 minutes.

Panel members are Mr. Hansson, Alan Purves (participating moderator), James Britton, and Julie Jensen. The discussion centers around three questions. What one means by "response to literature?" What sorts of responses English educators are interested in facilitating? and How can responses be measured? Topics include the validity of different models of appropriate responses, methods of doing research in this area, and the appropriateness of students' learning to respond to literature in the ways that have generally been measured.

Plenary Session Five:
"Research into Imagic Association and Cognitive Interpretations."
By HARRY BROUDY. 25 minutes.

Panel members are Mr. Broudy, Stanley Kegler (participating moderator), Martha King, and Bernard Shapiro. The discussion centers around the purposes and justification of formal schooling, the possibilities for measuring the effects of formal schooling, and the status of the knowledge with which education or any other discipline operates. Topics include the current challenge to traditional schooling, methods of measuring the impact of schooling, and the task facing those who would hope to justify formal schooling in our present society.

Plenary Session Six:
"Computer Applications—A Dynamic Medium for Creative Thought."
By ALAN KAY. 50 minutes.

Panel members are Mr. Kay, Thomas Barrett (participating moderator), Peter Rosenbaum, and Jaap Tuinman. The discussion centers around a new computer-like device called the Dynabook, the place of such devices in the schools, and what has been learned from experimentation with computer-assisted instruction. Topics include the sort of learning that can occur from students' manipulating such devices, the impact of technology on education, and the

possibilities of structuring peer teaching situations in such a way that gains similar to those achieved with computer-assisted instruction are realized.

Plenary Session Seven:
General Discussion with Consultants. 50 minutes.

Panel members are Richard Braddock (participating moderator), Harry Broudy, Jeremy Finn, Gunnar Hansson, Carol Talbert, and Sloan Williams. In this final plenary session, the consultants and others reflect on the Seminar and the ideas generated. In some cases, ideas presented earlier are reinforced and clarified. In others, ideas that were accepted earlier are questioned and perhaps modified. Topics include the political realities of getting research findings implemented and the question of whether adult or child models of competence ought to guide research.

A Conversation on the Seminar. 30 minutes.

In the final meeting of the Seminar, Dwight Burton and James Squire talk about the conference—what happened, what did not happen, how this conference differed from previous ones, and what future conferences of this sort may be like. Topics include the relevance of research, the need for evaluation and ranking of research, the training of graduate students as researchers, the kinds of research that need to be done, and the proper focus for research on education.

By way of conclusion, I will briefly mention some possible uses of the tapes. One possibility is for the entire set of tapes to be used to recreate the conference for graduate seminars in research in English Education. After being told about the Seminar and perhaps seeing the composite tape, students would become Seminar participants. They would read a paper, attend the plenary session on the paper via the tapes, and then form their own working parties to discuss implications of the paper and the plenary session for their particular research interests. This process would be repeated for each of the papers. Then students could assess their own conclusions on the Seminar against the comments made in the *General Discussion with Consul-*

tants and the Burton and Squire *Conversation on the Seminar.*

A second possibility is that several of the tapes could serve as a focus for a two or three day conference on research in English Education. Here, the coordinator of the conference would select those tapes he considered most relevant to the participants and run the conference somewhat like the seminar described above.

A third possibility is that one or more of the tapes and papers might be useful in various courses. To give just two examples, *Plenary Session Four*: *"Some Types of Research on Response to Literature"* should be particularly useful in courses on the teaching of literature. And *Plenary Session Five: Research into Imagic Association and Cognitive Interpretations"* ought to prove extremely provocative to any one interested in the future of the public schools.

I could go on, for every tape and various conjunctions of them offer interesting possibilities. Let me conclude by echoing again Rosenbaum's comment on the Seminar. For those of us who attended, the Seminar *was* an experience. We hope the video tapes will afford others the opportunity to share that experience.